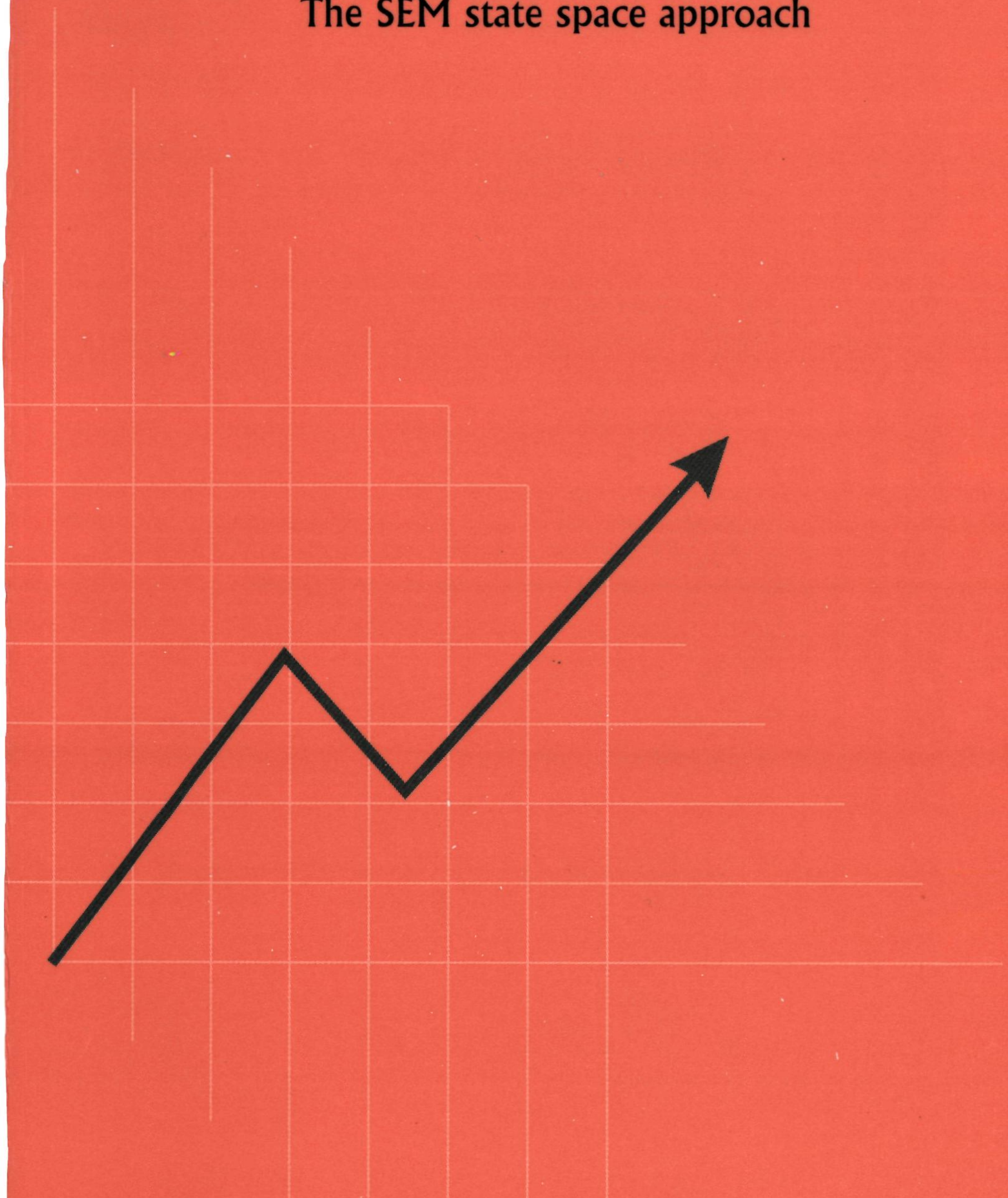
Constructing monitoring systems in the behavioral sciences

The SEM state space approach



Robert A.R.G. Jansen

Constructing monitoring systems in the behavioral sciences

The SEM state space approach

Uitgeverij KU Nijmegen
Postbus 9102
6500 HC Nijmegen
Tel.: 024 - 361 20 73

ISBN 90 373 0370 6
NUGI 724

Constructing monitoring systems in the behavioral sciences

The SEM state space approach

een wetenschappelijke proeve op het gebied
van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen in het
openbaar te verdedigen op maandag 12 mei 1997,
des namiddags om 1.30 uur precies

door

Robert Albertus Rudolf Gerardus Jansen
geboren 31 januari 1968
te Cuijk

University Press Nijmegen

Promotor: prof. dr. E.E.J. De Bruyn
Copromotores: dr. J.H.L. Oud
dr. J.F.J. van Leeuwe

Leden van de manuscriptcommissie:
prof. dr. A.J.A. Felling (voorzitter)
dr. C.A.J. Aarnoutse
prof. dr. M.P.F. Berger (RUL)
dr. R.N. Eisinga
dr. D.M.A. Haughton (Bentley College, Waltham, MA, USA)

TO MY FATHER

Contents

List of Figures

List of Tables

Preface

The general aim of this thesis is to contribute to the development of the design and application of monitoring systems within the field of education. The first chapter introduces the concepts of monitoring and monitoring systems. Chapters 2-6 treat methodological and statistical problems of interest. The last chapter contains a discussion of the various chapters, especially as regards the implications for the construction of monitoring systems within the field of primary school education.

Chapters 2-6 have originally been written in the format of journal or proceedings articles and thus are self-contained. It has led to a considerable overlap, however, because in each chapter the formulation and parameter estimation of the SEM[1] state space model is, to some extent, discussed. Furthermore, decisions concerning the references cited in the text had to be taken. Although a reference might have been cited in a previous chapter, in each chapter it is fully cited the first time it is referred to. In some cases, without mentioning, a reference is made to an article which is also part of the thesis (e.g. to chapter 2).

In *chapter 1* a general introduction of monitoring systems is given. Because of the rich variety of monitoring systems encountered in practice the subject is placed within a broad context first. Key concepts are discussed, objectives, and other aspects that guide the process of monitoring in practice. Second, monitoring systems are placed within the context of educational assessment. The SEM state space model as a central approach to the monitoring of educational growth in primary school education is discussed as well as two other approaches. The first chapter contributes to the central theme of the thesis in two ways. It provides general background information on the rationale of monitoring and monitoring systems, and it marks out the relevant context of the methodological and statistical problems discussed in chapters 2-6.

Chapter 2 treats the problem of missing data in panel data sets. Although missings may occur in all types of social science research, panel research is especially prone to produce incomplete data. Missing data often imply loss of information, especially in case a researcher performs analyses on the basis of subjects having complete data only. After a number of repeated measurements attrition often results in a reduction of the sample size to less than 50% of the original size. Missingness often relates to the values of the outcome variables. For example, pupils whose achievements are low tend to have a higher nonresponse to tests at later occasions just because of their low achievement levels. As a result, data may be systematically missing, which in general leads to severe problems in obtaining correct population parameter estimates. Several missing data mechanisms are distinguished, which, depending on the missing data procedure used, can be defined

[1] SEM is short for Structural Equation Modeling commonly known in social science methodology

to be ignorable or nonignorable.

A missing data procedure is proposed for the SEM analysis of panel data sets, using the EM (Expectation-Maximization) algorithm in conjunction with the Kalman smoother for computing maximum likelihood estimates of longitudinal SEM models from varying missing data patterns. It is a model-based procedure assuming certain relationships between the variables of interest. The EM algorithm allows the data to be missing at random (MAR) and has several advantages compared to so-called 'ad-hoc' procedures.

Chapter 3 gives an extension of the longitudinal SEM model in the treatment of missing data. It is shown how the missing data procedure can be utilized in the construction of the pupil monitoring system LISKAL. Instead of a 'zero means' SEM a 'structured means' SEM is defined, implying that the state space model (SSM)[2] is to be formulated in terms of latent and observed means processes as well. Also the Kalman filter part of the Kalman smoother is shown to accommodate for the latent and observed means processes. As a result, individual latent developmental curves can be estimated on an absolute scale. A pupil's level of achievement at a certain time can be compared to previous or later times in terms of absolute growth or decay. At the same time, however, his position can be assessed relatively in comparison to the absolute latent mean growth curve. In chapter 3 the term SEM instead of LISREL is used because the missing data procedure can also be performed by making use of any other SEM program than LISREL.

In *chapter 4* general nonstandard linear and nonlinear constraints are employed in longitudinal SEM modeling of panel data. First, nonstandard constraints are applied for the modeling of first-order and second-order stationary processes. General matrix algebraic expressions are derived and applied to constrain the first- and second-order moments of the latent and observed variables of interest. The implications of stationarity for the time dependence of the model parameters are clarified. Second, nonstandard constraints are applied in the modeling of growth on the basis of the overlapping cohort design (OCD) or accelerated longitudinal design. Such a design is important in an efficient construction of monitoring systems and has been extensively used in the construction of LISKAL. By means of nonstandard constraints it can be tested whether the partially overlapping cohorts have common model implied characteristics in the form of latent mean trajectories and latent covariance functions, which subsequently can be estimated and used in the monitoring system. On the basis of common latent means and latent variances-covariances developmental curves can be extracted that cover the entire time span of interest.

In *chapter 5* different models are shown to be special cases of the basic SSM and to be translatable into SEM. The Kalman filter and Kalman smoother are

[2] SSM is short for State Space Model.

applied to the structured means SEM model and the state-trait model. The SEM formulation and parameter estimation of the SSM with input-effects is discussed in detail. A comparison of the Kalman filter with two well-known cross-sectional factor score estimators is made. An important question to be answered is which of the two cross-sectional estimators should be used for initializing the Kalman filter as regards minimum variance and unbiasedness. The relationship of the Kalman smoother to the 'overall' regression estimator is shown as well as the t_0-conditional unbiasedness of the smoother. Finally, the problems of initialization of the Kalman filter on the basis of the state-trait model are solved by means of the Bartlett estimator.

Chapter 6 discusses maximum likelihood estimation of the continuous time linear stochastic SSM by means of SEM. Under rather general conditions, assuming the parameter matrices to be piecewise time-invariant or varying continuously over time according to a polynomial scheme, the so-called exact discrete model (EDM) is derived. As the parameters of the EDM are complicated nonlinear functions of the original parameters of the continuous time system, complex nonlinear relationships must be employed in SEM during the estimation procedure. This procedure is also known as the 'direct method' and represents an alternative to the heavily criticized 'indirect method' formerly employed in SEM.

Finally, in *chapter 7* the major findings of the thesis are restated. The relationships between the contents of the different chapters are emphasized. Special attention is given to the relevance of the different contributions for the construction of monitoring systems within the field of primary education. Furthermore, critical comments are given on some of the results as well as a number of suggestions for future research.

Chapter 1

Monitoring systems

1.1 Introduction

There is a growing interest in the development and application of monitoring systems in various fields like education (Plomp, Huijsman & Kluyfhout, 1992, Scheerens, Stoel, Vermeulen & Pelgrum, 1988, Tuijnman & Postlethwaite, 1994), project management (Casley & Kumar, 1988, Pelgrum, 1990, Van der Putte, 1991), and the study of environmental processes (Gosovic, 1992) In the field of the medical sciences monitoring systems are well known, as for example in quality improvement programs of health care (e g Jencks, 1994), and in the monitoring of physiological processes in cardiac surgery (e g Marangoni, 1994, Vitacca & Clini, 1994)

Several reasons can be given for the increase in the use of monitoring systems For example, the wish of policy makers to gain control of a variety of social and global processes The growing complexity of modern society and the continuing changing conditions of life ask for elaborate methodologies to make monitoring and controlling activities effective At the same time, developments in the field of data collection and data processing, and in computer technology, provide the means by which these activities can be performed To give an impression of monitoring and monitoring systems, three examples from very different fields are presented first

1.1.1 Examples

Example 1

Respiratory monitoring in an intermediate intensive unit. (source Vitacca, M & Clini, E 1994)
The major goal of monitoring is continuous recording of indices that enhance our understanding of the underlying pathophysiology, in order to improve diagnosis and guide management, and identify trends that assist in assessing the therapeutic response and predicting prognosis Nowadays, technology has made it ▷▷▷

possible to automatically sense and display a wide variety of physiological indices An ideal monitoring system should be pertinent to patient management, propose interpretable data, show high technical accuracy, high sensitivity, good reproducibility, be practical to use
The international literature, our personal experience, and cost considerations have proposed the following monitoring standards as the best for a noninvasive respiratory intermediate intensive care unit (RIICU) 1) mandatory indices respiratory rate, oxygen saturation, haemogasanalysis, tidal volume, minute ventilation, maximum voluntary ventilation, forced expiratory volume in one second, forced vital capacity, vital capacity, maximal inspiratory pressure, heart rate and blood pressure, 2) second choice indices capnometry, respiratory inductive plethysmography, transcutaneous monitoring of gases, haemodynamic monitoring, mechanics data by means of an oesophageal balloon, and central drive
Pulmonary monitoring devices shorten the time for patients who remain on mechanical ventilators, a reduction both in the risk of associated complications and the costs involved is a natural consequence Continuous monitoring of significant physiological indices has the potential for predicting a critical event, and providing an opportunity for the institution of lifesaving measures
In conclusion, a RIICU where it is possible to admit patients from various other intensive care units (ICUs) or those requiring preventive care until aggressive therapy is essential, must be equipped with adequate, noninvasive and less expensive bedside respiratory monitoring devices ○ ○ ○

Example 2

A system for Monitoring and Evaluation & Management Information (MEMIS); an integrated system for planning purposes in developing countries (source Van Tilburg, De Haan & Giesberts, 1995)
MEMIS is a combined system of **monitoring & evaluation** and **management information**, The system has in the first place been developed in support of a large-scale integrated rural development project in Zaire, involving many different fields of activity, like medical care, agricultural productions and processing, education and emancipation, water facilitation, road maintenance, garage and transport The project area comprised of a territory as large as Benelux, covering a target-group of about half a million people The initial aim of MEMIS was to be a tool for project management enabling it to perceive the successes and failures of the project as a whole, as well as its individual working-units or institutions At the same time MEMIS makes regular information on the progress of the project, including related costs, available for the donor organizations The system ▷ ▷ ▷

makes use of a specifically designed software package for storage and tabulation of data for decision-making purposes In other words, MEMIS is a system complying with the following sequence of activities (storage and presentation of data), and again manual activities (analysis and decision taking)
MEMIS, as a monitoring and management tool for project management as well as for donors, can be used to monitor and, based on a consensus between management and donor, possibly adjust the project This applies to large and structurally complex projects, but to more modest ones as well

MEMIS as a system refers to the overall package, covering all the mentioned phases Moreover, the MEMIS-program is an integrated part of the system and its mathematical core This is a tailor-made software package in which data are imported and stored, and which makes all required tables and questionnaires The MEMIS-program is an extremely user friendly package, protected as much as possible against mistakes and coincidentally made mis-tappings ○ ○ ○

Example 3

(source Van der Putte, 1991)
ARTEMIS (African Real Time Environmental Monitoring System), developed by the Food and Agriculture Organization of the United Nations (FAO) in collaboration with NASA (North America Space Agency), ESA (European Space Agency) and NOAA (National Ocean and Atmospheric Administration), is a system to monitor precipitation and vegetation conditions on a continental scale Its development was made possible by the technological developments in the field of remote sensing which made operational use of satellite imagery feasible, and the development of computers, that permitted the processing of vast quantities of data using complex formulas It was initiated as a reaction to the recurring problems in food production and the recurrent locust plagues in large parts of the African continent
The ARTEMIS monitoring system is fed with data from Meteosat (a geostationary weather satellite from ESA) and by NOAA polar orbiting satellites The Meteosat data are directly received by the Meteosat Primary Data User Station (PDUS) NOAA satellite data are received on tape Other data sources include maps and observations from weather stations in the region The processing of the data acquired by ARTEMIS is orientated towards the two main products of the system ▷ ▷ ▷

(1) rainfall estimation per dekad (10 day period) and month using hourly data from Meteosat,

(2) vegetation monitoring based on NOAA data using the Normalized Difference Vegetation Index (NDVI).

ARTEMIS provides maps on a continental scale and for sub-areas as well as statistical data on geographic locations. The output comprises (Hielkema and Howard, 1986):

(1) raw data of both Meteosat and NOAA/AVHRR (Advanced Very High Resolution Radiometer),

(2) NDVI, estimated rainfall and estimated rain-days per ten day period and per month,

(3) monthly estimated rainfall anomaly map,

(4) a ten day potential locust breeding activity factor.

The output should allow users to identify eventual problems to be expected with regards to agricultural production and to assess breeding conditions for the desert locust. The former should support the prediction of food deficiencies, and hence would support the management of food stocks. The assessment of the breeding conditions for the desert locust supports the management of locust control programs.
The system is still under development as are the applications of the systems output. In the initial phase of systems design the following has been said about the system users (Hielkema and Howard, 1986):
"The potential users group of the ARTEMIS information products is very heterogeneous and consists of both operational users and users of the data for research purposes. So far the following have shown interest as users:

* *FAO Global Information and Early Warning System (GIEWS)*

* *FAO Emergency Center for Locust Operations (ECLO)*

* *FAO Office for special relief Operations (OSRO)*

* *FAO Agro-meteorology Group*

* *FAO Land and Water Development, Forestry and Fisheries development/monitoring programs* ▷ ▷ ▷

* *Regional/national Early Warning Projects*
* *Regional/national Desert Locust Organizations*
* *Other UN Agencies, e.g WFP, UNEP, UNESCO, WMO."* ○○○

As the examples 1, 2 and 3 show, monitoring systems can be employed for various reasons but also differ substantially in the way the monitoring function is implemented. In general, monitoring relates to a number of concepts which characterize the process of monitoring.

1.2 Concepts and definitions

As several authors indicate, there does not exist a generally accepted and consistent theoretical basis of monitoring (Van de Putte, 1991; Van Tilburg & De Haan, 1995). Much of the literature on the subject describes experiences with monitoring systems in practical situations. As the context and the objective(s) of monitoring give a monitoring system its meaning (Casley & Kumar, 1988), these can be very diverse. Monitoring involves repeated assessment and the repeated collection of data, possibly consisting of independent replications in the cross-sectional dimension (Hsiao, 1986). It supposes a time frame in which the process takes place, and utilizes objective criteria and standardized procedures. A monitoring system is user-oriented, providing a user or a number of users with information. Users may be individuals, the public, organizations or a number of individuals within an organization, depending on the type of inquiry and the objective(s) of monitoring.

Monitoring is related to concepts like assessment, evaluation, intervention, control, and the systems perspective. Within the behavioral sciences a feasible definition of monitoring can be given. Monitoring commonly refers to systematic and regular procedures for the repeated collection and interpretation of assessment data of important aspects of the subject under study. It is not necessarily restricted to outcome variables, but can also involve contextual information and measures of inputs and processes (Husén & Tuijnman, 1994; Scheerens et al., 1988).

Whereas assessment refers to techniques of determining outcomes, either by subjective judgments or by means of standardized objective tests, evaluation involves the making of a judgment as to the achievements of the outcomes, and is therefore subjective (Van der Putte, 1991, p. 33). Just like monitoring, evaluation has a different meaning in different contexts and can be performed in several ways (e.g. Casley & Kumar, 1987, 1988; Husén & Tuijnman 1994). For example, it can provide feedback on an intervention that has been completed (ex-post evaluation), or it can be an activity during the monitoring of an intervention (ongoing evaluation). Evaluation can take place from a comparative point of view. For example, outcomes of different learning programs may be evaluated in terms of students' performances. It can also be performed in terms of standards or a predefined set of goals. A standard, for example, is applied in case a mean score of a well-defined student population on a set of test items is used to assess how well individual students perform over time.

Evaluation and monitoring are related concepts. Monitoring, at various levels of detail, provides information that ultimately defines the focus of the evaluation activity. For this reason evaluation is sometimes defined as a part of the monitoring function (Pelgrim, 1990, p. 8; Plomp et al., 1992; Van de Putte, 1991). Note, that although evaluation is liable to subjective judgments, objective criteria can be employed in the evaluation process.

The general objective of a monitoring system is to provide a user or a number of users with several sources of information with regard to the process being investigated. Information may give a cause to intervene (Casley & Kumar, 1988), that is, to take some steps or measures to influence or to control future behavior or developments. Monitoring based intervention is an important means of controlling both behavioral and nonbehavioral processes.

A basic model for conceptualizing the various aspects of a monitoring system is derived from the systems perspective (Pelgrum, 1990; Plomp et al., 1992; Scheerens, et al., 1988). A system is a model of reality which consists of a set of entities that evolve and interact as time progresses. The interaction between the system and its surrounding is realized via inputs and outputs. There exists a mapping between the inputs and outputs which defines the process or throughput of the system. A system can be characterized by a set of goals. These often determine the feedback mechanism within the monitoring system. Discrepancies between system goals and system outputs ask for adjustments to be made in terms of inputs or throughputs. Providing feedback is sometimes referred to as the main function of monitoring (Pelgrum, 1990, p. 7). The systems perspective is very useful even in complex surroundings in which many variables interact and influence the outcomes of a process. Moreover, it allows the use of mathematical representations. Mathematical representations of dynamic phenomena are flexible and can involve highly complex systems (Luenberger, 1979; Willems, 1991).

Two other concepts, which often are crucial in monitoring, are to be mentioned also. First, a monitoring system may be utilized for the objective of forecasting. It presupposes a model in which the mechanisms, which are thought to underlie the processes of interest. are more or less formalized. In case the model is fed by repeated measurement data, forecasts are obtainable on the basis of the knowledge already contained in the system. Second, monitoring techniques may be employed for the objective of problem signaling and diagnosing. For example, a monitoring system points out (or predicts) a problem at a certain time because a standard or set of standards has not been satisfied. In case the system is equipped with diagnostic tools, problems may additionally be analyzed for the purpose of problem solving and the facilitation of remedial action (e.g. Aarnoutse, van Leeuwe, Oud, Voeten & van Kan, 1996a; Casley & Kumar, 1988).

As the concepts above indeed characterize the process of monitoring, there are numerous other aspects which determine the scope of monitoring and the final shape of a monitoring system. The three examples above reflect the potentially rich variety of monitoring systems. Decisive, however, are the objectives of monitoring, the focus of the monitoring function, and the organizational setting in which monitoring is to be performed.

1.3 The scope of monitoring

1.3.1 Objectives of monitoring

Within the context of a monitoring system the perspective of monitoring and its objectives become defined. The general objective of monitoring relates to 'quality' and 'control' of the process(es) being monitored (e.g. Casley & Kumar, 1988; Scheerens et al., 1988). It implies that monitoring relates to policy making processes. In fact, monitoring is seen as a management tool in a broad sense (e.g. in project management, Van der Putte, 1991, but also in guiding management in respiratory monitoring, Vitacca & Clini, 1994). It can be employed for the preparation of policies and the evaluation or implementation of policy measures.

The general objective of what is termed 'quality' and 'control' gets its precise meaning within its context. In example 1 (see page 1), the major objective is to enhance the understanding of the underlying pathopsysiology, which is further decomposed into four other objectives (i.e. improve diagnosis, guide management, identify trends in assessing therapeutic response, and predicting prognoses). As each context of monitoring has its specific characteristics and requirements, a complete overview of objectives can impossibly be given (e.g. Van der Putte, 1991; Van Tilburg & De Haan, 1995, pp. 9-11).

As the objectives of monitoring serve as guidelines in determining the foci of the monitoring function, these also determine how monitoring is to be performed in practice. Monitoring practices, however, also depend on the nature of the information, the setting in which this information is to be obtained, and the state of knowledge.

1.3.2 Foci of monitoring practices

Monitoring may be directed at the inputs, processes or outputs (Plomp et al., 1992; Scheerens et al., 1988). Input monitoring concerns the repeated assessment of the input of a process (e.g. individual or group background, or financial and material resources of an organizational unit). Process monitoring, which obviously relates to the processes of interest, is performed to assure that these contribute to the (expected) outcomes. It can involve various aspects, such as characteristics of the organization (e.g. decision-making procedures) or the curriculum in case of monitoring the quality of education. Finally, output monitoring is directed at the outcomes of a process, such as the assessment of project achievements, also called performance monitoring (e.g. in MEMIS, see example 2 on page 2), or the assessment of project efficiency (e.g. 'Have the means been sufficient to realize the goals?').

In project management monitoring often is a control procedure for the implementation of interventions, also called implementation monitoring. It is associated with short-term objectives (Van der Putte, 1991) and directly supports the decision-making process. Strategic monitoring is based on long-term objectives. It has been developed in the context of regional and national planning and aims at sustaining strategic management. It operates in highly complex environments and provides a framework in which other monitoring systems, at lower levels of management, function. Its immediate objectives may not be clear but are to be defined at the lower levels. For example, in implementing regional economic developmental programs (Casley & Kumar, 1988).

A monitoring system is descriptive in case there are not any assumptions with regard to the specific effects between the variables of interest. A causal model can also underlie the monitoring system, however, aiming at a causal explanation of the outcomes of a process. Whether or not causal schemes are involved depends on the state of knowledge. In modeling variables by means of causal relationships an effort is made to enhance the understanding of the process under study.

The registration or assessment of outcomes may be the only function of monitoring, meaning that no actions or interventions with respect to the processes being monitored are undertaken. For example, the precipitation conditions monitored by ARTEMIS (see example 3 on page 3) cannot be controlled by any means. However, it provides insight into the changing conditions of vegetation, and because these conditions relate to agricultural production and vica versa, the monitoring information is crucial in supporting management.

Monitoring information is usually obtained by making use of formal standardized procedures of data collection, data processing and assessment. However, sometimes an informal approach is advocated; if information is highly unstructured or if it is unclear how and from which persons or settings it is obtainable. In these cases data collection and data processing procedures are less structured and require a pragmatic approach, as for example in day-to-day management (e.g. Plomp et al., 1992). Nonstructured informally obtained data can provide useful insights to supplement the formal system of monitoring (Casley & Kumar, 1988).

Monitoring practices obviously are very diverse. A monitoring system, however, can only be successful if its tailored to the requirements of the organizational setting in which it is meant to function.

1.3.3 The organizational setting of monitoring

Monitoring systems are naturally placed within an organizational structure and often function within different segments of an organization. Several authors (Casley & Kumar, 1988; Van der Putte, 1991) state that at each segment the monitoring function and the information which results from it should be tuned to the specific requirements of that organizational unit. That is, the information should be indicative for the specific functions and responsibilities of the persons who make use of it. Only if the flow of information is integrated within the segments of the organization, and only if the objectives of monitoring within these segments are clearly defined, a monitoring system can be successful (Van der Putte, 1991).

Information at higher levels of the hierarchy normally ask for higher levels of data aggregation (Casley & Kumar, 1988). In education, for example, a teacher is primarily interested in the students' performances, whereas the school principal is interested in the average school performances (averaging over the students' achievements attending the school), as to make between school comparisons possible (e.g. Scheerens et al., 1988).

The organizational surrounding of a monitoring function thus provides information about its objectives, actual content, and design of the monitoring system. In general, the selection of information in monitoring is based on the knowledge of the process(es) of interest and, in fact, defines the critical variables of the monitoring system. The measurability of the variables or proxies also determine which variables are to be used (Van der Putte, 1991; Scheerens et al., 1988). Measurability can be limited by practical conditions. At schools, for example, it is usually too demanding to perform extensive measurement procedures because of the size and relevance of the learning program (Gillijns, 1991). Additionally, the choice of the variables and the nature of the measurements (quantitative versus qualitative) relate to the choice of data processing techniques (Casley & Kumar, 1988; Van der Putte, 1991).

A final remark considers the time schedule of monitoring and the actual construction and implementation of a monitoring system in relation to the organizational setting. The construction of a monitoring system involves a number of activities, such as the development and testing of measurement instruments, the planning of sampling and data collection, and the planning of data processing.

These and other activities have to be geared to one another as well as to the relevant organizational setting in which the monitoring system is going to operate.

1.4 Monitoring educational growth

In the past decades much effort, national as well as international, has been given to questions about the quality of education. In many countries this has led to programs for the implementation of national assessment studies of educational progress (Husén & Tuijnman, 1994; Pelgrum, 1990). In the Netherlands the advisory council of primary education (ARBO, 1988) has stressed the importance of a national pupil monitoring system and gave the initial impetus for defining its objectives and content. Whereas the objectives and content have been more or less clarified, several methodological approaches exist to design a national monitoring system (e.g. Gillijns, 1991).

This study focuses on the development and application of methodological and statistical tools within the SEM state space approach for the analysis of behavioral processes and the construction of monitoring systems. Although the SEM state space approach is generally applicable to the modeling of processes, it has been mainly employed for the development of the pupil monitoring system LISKAL (Aarnoutse et al., 1996a). In fact, a number of problems were raised at schools while working with the LISKAL system. These problems, which have been addressed in this study, are methodological in nature and are especially relevant within behavioral science applications of monitoring. The study is restricted to behavioral science processes and more specifically to the construction of monitoring systems in primary school education. The relevant context is the primary school in which monitoring aims at registering individual learning progresses on the contents of the curriculum.

In stating the research problems and introducing the SEM state space model, a description of other methodological approaches to the monitoring of educational growth is appropriate. In fact, all of these systems have to provide answers to the same type of research questions, although these have been solved differently. First, the SEM state space model as a general approach to the modeling of dynamic processes is introduced. Second, two other approaches to systematically monitoring educational growth at the primary school level are discussed. The first approach is only addressed for the sake of completeness but is not extensively discussed because it does not involve any elaborate statistical test theory. The second approach utilizes Item Response Theory (IRT), which has a rich tradition in the field of psychometrics (Hambleton & Swaminathan, 1985).

1.4.1 The SEM state space approach

The SEM state space model combines two approaches from different fields. Whereas the SSM results from system and control theory and has its roots in the physical sciences, the SEM model originates from disciplines within the social sciences. Because of its generality and flexibility the combined approach of the SEM and the SSM model seems promising, especially for behavioral science applications (MacCallum & Ashby, 1986). Monitoring systems based on this approach are one of its recent results.

The SSM covers a broad class of dynamic models (e.g. Oud, 1996; Singer, 1992) and is widely used for the analysis of stochastic processes in engineering, econometrics, time-series analysis and other related areas (Caines, 1988; Jazwinski, 1970). Application of the SSM allows the use of optimal filtering and smoothing techniques as well as procedures for optimal control (Jazwinski, 1970; Rauch, Tung & Striebel, 1965). The linear stochastic discrete time SSM consist of a dynamic state equation and a measurement or output equation. Because these equations and its properties are extensively discussed later on, these are not presented here. Instead, a number of essential characteristics of state space modeling are discussed.

As the term indicates, system theory employs a systems perspective. A dynamical system consists of inputs, outputs and states, often functions of time (Caines, 1988; Zadeh & Desoer, 1963). Its relationships, possibly functions of time as well, can be mathematically formalized to study the behavior of dynamic phenomena. Mathematical system theory, which involves the theory on difference and differential equations and linear algebra, provides the technical means for modeling and analyzing stochastic and non-stochastic dynamic processes. There exist many applications, which range from purely technical matters, such as in optimal control theory (e.g. satellite orbit estimation, Lewis, 1986), to sociological types of phenomena (e.g. reciprocal effects of interclass marriages and social structure, Luenberger, 1979).

A dynamic system is externally represented by a mathematical description of the input and output functions and its mutual relationships. An internal description involves the definition of the system's state or more generally, the system in state space form. The notion of 'state' is fundamental in system theory. It summarizes the system's past behavior which together with the future inputs determine all future states and system outputs. Whereas the state points to a condition of a measured quality at a certain time, space points to the set of all possible conditions of that state. In addition to the state space there exist an input and output space as well (see Willems, 1991). There are a number of properties which underlie the SSM (e.g. Caines, 1988, appendix 2) of which a few are mentioned here.

The non-anticipativity property holds that the state is non-anticipating, excluding effects going backward in time. It is also referred to as the causality principle and coincides with the generally held conception of processes evolving in time, namely, that the past influences or determines present and future states instead of the other way round. The state separation property, which is essential for defining the concept of state, holds that the state contains all the information on the past behavior of the system. If the state is completely known, all past information can be disregarded. It implies that effects between states exceeding more than one time lag are not allowed. In case this property is violated (e.g. models containing higher-order between-state effects such as autoregressive moving average models or ARMA), the model is easily reformulated as to satisfy this property, however (Caines, 1988, p. 111). Finally, the instantaneous-output-map property holds that the effects between the states and the outputs are not only unidirectional but also instantaneous. There exist many applications as well as variations of SSM modeling. For a general treatment, see for example Caines (1988) and Luenberger (1979) and the literature cited therein.

Structural equation models (SEMs) have been widely used in fields like psychometrics, econometrics, biometrics, and sociology. It is from these fields that many scholars contributed to its development (Bentler, 1980; Bollen, 1989; also Bollen & Long, 1993; MacCallum & Ashby, 1986). SEMs have become very general and flexible which explains its past and present popularity. Much of its development is indebted to the pioneering work of Karl Jöreskog (1967, 1969, 1973).

An important advantage of SEM, also with regard to the SSM, is that it combines confirmatory factor analysis models with simultaneous equation systems. The general SEM, however, includes a whole range of statistical techniques, such as regression analysis, path analysis, multiple indicator analysis, and panel data analysis. It allows measurement errors in both the exogenous and endogenous variables. It can handle latent and observed variable models with correlated errors, as well as metric and nonmetric data, and allows general linear and nonlinear constraints to be made.

In SEM a clear distinction is made between the measurement and structural model. The measurement model relates the observed variables to the unobserved latent ones by means of a set of parameters. The structural model specifies relations and effects, also known as parameters, between a number of latent or possibly observed variables. In fact, a whole range of models can be specified (Jöreskog & Sörbom, 1989) of which the model parameters are estimated in a SEM analysis.

Analyses proceed on the basis of the sample covariance or moment matrices of observations, assuming the sample consists of independent replications in the cross-sectional dimension (e.g. Hsiao, 1986). The fundamental hypothesis to be tested is that differences between the sample covariance/moment matrix and the covariance/moment matrix implied by the set of structural and measurement equations and its accompanying parameters are minimal (Bollen, 1989). If these differences are minimal, the model approximates the observed covariance/moment matrix very well. It yields a chi-squared (χ^2) value to evaluate the goodness-of-fit of the model to the data, and standard errors which reflect the sampling variability of each parameter estimate. Based on the χ^2-value, a number of fit indices can be employed for model selection and evaluation (for a treatment see Oud, Haughton & Jansen, 1996; Haugthon, Oud & Jansen, 1996).

The SEM model is a linear model. Several estimation procedures may be utilized (e.g. Jöreskog & Sörbom, 1989, p. 16-23; Neale, 1995, p. 54-59). A well-known and very often used procedure is the maximum likelihood estimation method. It maximizes the likelihood of the parameters given the data, and assumes the observed variables to be jointly multinormally distributed. It is also known to be quite robust against deviations from normality, however (e.g. Boomsma, 1983; Muthén & Kaplan, 1985).

As has been shown in a number of papers (MacCallum & Ashby, 1986; Oud, 1996; Oud, Van den Bercken & Essers, 1990) the SSM can be represented as a special case the SEM model. Because in SSM modeling a number of restrictions apply, the SEM model can be seen as the most general one. Because in the physical sciences the model parameters and states are mostly known in advance, SSMs often are constructed by engineers themselves. In the behavioral sciences, however, this is not the case. Procedures are necessary to estimate the model parameters and the unobserved states. The SEM model provides a general and flexible framework in which the modeling questions can be answered and parameter estimation can be performed. One of the most important results in linear system and control theory has been the introduction of the Kalman filter (Kalman, 1960; Kalman & Bucy, 1961) and Kalman smoother (Rauch, et al., 1965). These methods, which are generally applicable for SSMs, can be employed for the estimation of the unobserved states, and possibly for a small number of unknown parameters in the model. The SEM state space approach allows the application of the Kalman filter on the basis of the SEM model as to provide optimal estimates of the latent states. Also, because SEM models accommodate for time-varying and time-invariant model parameters, the SEM state space approach has substantial value in longitudinal behavioral science investigations.

1.4.2 The pupil monitoring system LISKAL

The 'Nijmegen pupil monitoring system LISKAL' is a computer based pupil monitoring system designed for primary school children from grade 3 up to grade 8. It allows teachers to monitor the pupils' achievements with regard to reading, language and arithmetic skills and accurately predicts future developments. Because of this, LISKAL is able to recognize learning problems in an early stage such that timely help can be provided when needed.

LISKAL employs a number of tests with regard to reading comprehension, spelling, decoding speed, arithmetic, and vocabulary (Aarnoutse et al., 1996a;

Aarnoutse, van Leeuwe, Voeten, van Kan & Oud, 1996b). It systematically registers and assesses the outcomes of pupils' test scores. It provides guidelines in examining the nature and cause of learning problems (diagnosing), as well as programs for solving these problems (remedies). Finally, the effects of remedies can be evaluated by comparing the predicted developments with the actual developments, that is, after the corrective measures have been taken and test scores have been obtained.

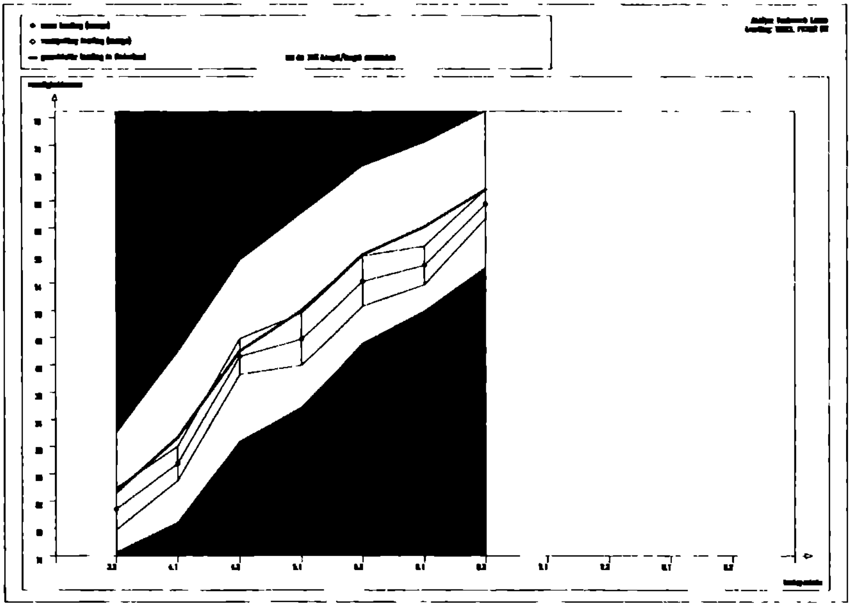
LISKAL employs several linear causal dynamic SSMs which take the form of overlapping cohort designs and are analyzed by means of SEM. It uses the Kalman filter for the estimation of individual latent developmental curves. Each of the three successive cohorts (from grade 3 up to grade 5) consists of a representative sample of the Dutch population of primary school pupils (Aarnoutse et al., 1996a). By now, pupils have been monitored for a period of three years, with two measurements each year. The cohorts provide information on the population mean development and factor score standard deviations over time with regard to the contents of the curriculum. The population mean information serves as a norm of reference in assessing pupils' individual or average (sub)groups achievements over time.

For each of the five skills, absolute and relative developmental curves can be estimated. Inter- and intraindividual comparisons as well as comparisons between various (sub)groups can be made. Standard errors of estimation indicate the preciseness of individual estimates. By means of the Kalman filter, predictions of the expected achievement levels are obtained. Also, in case test scores are missing, the Kalman filter provides estimates on individual achievements. If LISKAL is employed for the whole primary school period, and combined with additional background information, it can provide information about school effectiveness.

Figure 1.1 displays an example of a graph obtained by the LISKAL program. The fat line represents the population's absolute mean developmental curve of decoding speed. The white band consists of the area plus and minus one standard deviation from the mean representing 68% of the population. The areas outside the white band each consist of, respectively, the 16% highest and lowest scoring pupils in the population. Pupils of which the latent scores lie within the area below the white band are assumed to have learning problems. The developmental curve of one pupil is represented by the thin line with filled in circles. The lines around it represent the standard errors of estimation (i.e. plus and minus one standard error).

The population mean development (fat line) shows an absolute growth over time. Between time points 4.1 and 4.2 absolute mean growth is highest. The pupil's developmental curve displays absolute growth also. From time point 3.2 to 4.1, the pupil's level increases, but individual growth is less than the average growth in the population. From time point 4.1 to 4.2 the individual growth is higher than the average growth, although the pupil's level is still below average.

Fig. 1.1: Absolute latent individual and population mean development of decoding speed.



1.4.3 Other approaches

Pupil monitoring systems based on educational age

A number of pupil monitoring systems, such as the SAVU system (Melis & Sonsma, 1989), are based on educational age (EA)[1]. The concept of educational age has been developed in analogy with the concept of mental age (MA). Mental age originally was introduced by Binet in 1908 (Wechsler, 1974, p. 1). It was divided by chronological age (CA) by William Stern in 1912, and finally used in the famous intelligence quotient, IQ = (MA/CA)100, by Terman in the first version of the *Stanford-Binet* in 1916. Analogously, educational age is used for the computation of the educational quotient (EQ), linking educational age to the didactical age (DA, the number of months a pupil has been educated) as follows: EQ = (EA/DA)100. Because of the serious problems connected with the use of IQ all serious intelligence tests as, for example, the Wechsler scales changed to standard scores (Wechsler, 1974, p. 1). The same problems apply to educational

[1] In Dutch it is refered to as 'didactische leeftijdsequivalent' or DLE.

age and the educational quotient, as explained by Moelands, Mommers and Oud (1990), Oud and Mommers (1990). Three main points are:

1. In using the EA and EQ a linear mean raw score development must be assumed over time and the raw score standard deviation must be assumed to be invariant over time. If, in terms of standard scores, a pupil's position in the raw score distribution remains the same over time, but the standard deviation increases, the pupil's position above (below) the mean is wrongly concluded to increase (decrease) in terms of EA and EQ. Comparable problems occur in the case of a nonlinear mean raw score development.

2. On the basis of the mean raw scores in the reference group, linear inter- and extrapolations of the mean development are made. If real development is not linear these estimates, and the EAs and EQs based on them, are biased, especially in case of large intervals between the measurement time points.

3. Educational age does not take into account the errors of measurement known from classical test theory. Lack of standard errors leads to the interpretation of meaningless differences between pupils and time points.

A pupil monitoring system based on IRT

The Cito pupil monitoring system (Gillijns, 1994), which is more recent than the SAVU system, consists of a descriptive nonlinear measurement part utilizing IRT, and a causal dynamic explanatory part utilizing the state space model (SSM). The measurement and structural model are explicitly separated and estimated separately. The measurement model makes use of the One Parameter Logistic Model (OPLM; Verhelst & Eggen, 1992) for the estimation of the item parameters and the individual abilities over time (Eggen, Engelen & Kamphuis, 1991). The individual abilities or latent scores and its standard errors of estimation can be graphically displayed and allow to make inter- and intraindividual or (sub)group comparisons on an absolute scale (e.g. Gillijns, 1991).

An IRT model specifies a relationship between the observable test performance and the unobservable trait or ability which is assumed to underlie the performance on the test. The items of a test and the individual abilities are placed on the same scale. The ability estimate gives information about item and test performance and allows a content oriented interpretation of test results. Ability estimates are independent of the particular choice and number of items that have been taken. It makes that individual abilities which are estimated on the basis of different subsets of items are comparable. In some estimation procedures (conditional maximum likelihood procedures) no assumptions have to be made about the distribution of the latent ability for the estimation of the item parameters.

Although the characteristics of IRT make it a flexible approach in test taking, because in adaptive testing each pupil is given its own selection of test items (e.g. Eggen et al., 1996), it is based on much stronger assumptions than classical test models (Hambleton & Swaminathan, 1985). The IRT model assumes unidimensionality of the latent trait. In case a trait is known to be composed of several dimensions (e.g. reading comprehension), the assumption is violated and the IRT model cannot be applied (e.g. Traub, 1983). In using IRT models for assessing growth, the unidimensionality assumption has to be checked at each time point. For making sure that the same latent trait is being measured over time, item parameters must be assumed to be constant over the entire time range. In IRT models it is also assumed that the individual responses to items in a test are statistically independent. This assumption of local independence states that the probability of any pattern of item scores of an individual is just the product of the probability of the occurrence of scores on each test item given a fixed ability level. The local independence and unidimensionality assumptions are equivalent (Hambleton & Swaminathan, 1985, p. 24). In case an IRT model satisfies these assumptions the advantages of this approach can be fully gained. However, the assumptions of the model are probably reasonable only if there is a close fit of the IRT model and the test data of interest.

The structural model utilizes the SSM and the Kalman filter for the estimation of individual growth curves. It consists of a causal explanatory scheme possibly involving several domains of the curriculum. Instead of a linear measurement model, however, it employs the nonlinear IRT model. As in standard Kalman filtering the conditional expectations of the state and state covariance matrix are derived under the assumption of a linear measurement model, the nonlinearities in the Cito approach lead to integrals which have no closed form and have to be evaluated numerically. Furthermore, the problem of estimation of the model parameters has to be solved, because standard SEM programs cannot be used (Kamphuis, 1992).

In studying growth by means of the SEM state space model, the latent variables have to keep the same content over time. In the LISKAL pupil monitoring system this requires the use of congeneric measurement instruments. Congeneric instruments measure the same underlying variables, meaning that the latent variables underlying the observed ones correlate one (Jöreskog, 1974). Although this imposes rather strong requirements for the instruments involved, the measurements of the same latent variable may be in different observed scale units and observed scale origins, with different reliabilities and even with a different number of observed variables. The congenericity concept is quite different from the unidimensionality assumption in IRT. It applies to whole instruments or tests instead of single test items, and therefore is far less restrictive. Furthermore, while in IRT individual items have to be unidimensional and have to measure the same trait over the entire time range, in LISKAL each instrument only needs to be congeneric at the time points it is applied to, often being not more than two consecutive time points. Finally, no local independence is required for the instruments applied at the same point in time. In fact, in the SSM, the measurement errors of instruments taken at the same point in time may correlate.

Chapter 2

Longitudinal LISREL model estimation from incomplete data using the EM algorithm and the Kalman smoother[1]

Abstract

Longitudinal data sets with the structure T(time points) $\times$ N(subjects) are often incomplete because of data missing for certain subjects at certain time points. The EM algorithm is applied in conjunction with the Kalman smoother for computing maximum likelihood estimates of longitudinal LISREL models from varying missing data patterns. The iterative procedure uses the LISREL program in the M-step and the Kalman smoother in the E-step. The application of the method is illustrated by simulating missing data on a data set from educational research.

2.1 Introduction

The occurrence of missing data is a general problem for both cross-sectional and longitudinal research. Dempster, Laird and Rubin (1977) first introduced the EM algorithm, which is a statistically well-founded and broadly applicable algorithm for computing maximum likelihood estimates from incomplete data. The EM algorithm has been employed in maximum likelihood estimation of a wide range of models (Little & Rubin, 1987). Shumway and Stoffer (1982) used the EM algorithm in conjunction with the Kalman smoother as an approach to parameter estimation, smoothing and forecasting for stationary time series with missing observations. A similar approach has been followed by Singer (1990, 1992, 1993).

[1] Reprint of Jansen, R A R G , & Oud, J H L (1995) Longitudinal LISREL model estimation from incomplete data using the EM algorithm and the Kalman smoother *Statistica Neerlandica, 49*, 362-377 Received May 1994 Revised May 1995

Shumway and Stoffer's procedure, however, is restricted to the $N = 1$ case (one single time series). Many data sets, especially in the behavioral sciences, have a $T \times N$ structure with $N > 1$. Further, in time series analysis, the number of time points T must be large to obtain estimates with reasonably low variances. This causes problems in fields with typically small numbers of repeated measures like the behavioral sciences. In addition, Shumway and Stoffer's time series model is stationary, which is a rather unrealistic restriction for many problems in the study of development (Oud, van Leeuwe, & Jansen, 1993). In the procedure proposed here for $T \times N$ structured data, a general non-stationary model is used and there are no requirements concerning the number of time points T. The model is the longitudinal LISREL model as derived from the discrete-time non-stationary (time-varying) linear stochastic state space model (SSM).

The proposed EM procedure uses the LISREL program in the M-step and the Kalman smoother in the E-step. The application of the method is illustrated by simulating missing data on a data set from educational research. For the computations the computer program LISMIS is developed. This repeatedly runs the LISREL program and computes Kalman smoother estimates between successive LISREL runs.

2.2 LISREL and state space modeling

The LISREL model to be estimated is formulated in terms of the SSM to make the Kalman smoother accessible for the proposed missing data procedure. Although this imposes certain restrictions on the LISREL model because of the causality principle inherent in state space modeling, it does not imply any substantial reduction in generality (Oud et al., 1993). The causality principle only requires the state to be nonanticipating, excluding effects going backward in time. The SSM consists of two equations: the dynamic part or state equation (Equation 2.1), which describes the dependence of the latent state variables in $\mathbf{x}_t$ on their lagged values in $\mathbf{x}_{t-1}$ and the static part or output equation, which connects the latent state variables to the observables in $\mathbf{y}_t$ (Equation 2.2):

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{w}_{t-1} \quad \text{with} \quad cov(\mathbf{w}_{t-1}) = \mathbf{Q}_{t-1}\ , \tag{2.1}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t \quad \text{with} \quad cov(\mathbf{v}_t) = \mathbf{R}_t\ . \tag{2.2}$$

The state transition matrix $\mathbf{A}_{t-1}$ in Equation 2.1 contains the autoregressive and cross-lagged effects between the state variables at successive discrete time points t and $t-1$: $t, t-1 \in \{t_0, t_0+1, \ldots, t_0+T-1\}$ for integers t_0 and $T \geq 2$, with t_0 the initial time point and T the total number of time points considered. The output or measurement equation (Equation 2.2) is equivalent to the factor model equation in factor analysis with $\mathbf{C}_t$ the factor pattern matrix.

Instead of Equation 2.1, many econometric and social science models choose a so-called structural equation, which has $\mathbf{x}_t$ at its right-hand side as well as its left-hand side. Before applying the Kalman smoother, however, such a structural equation can be reduced to Equation 2.1 (Oud, van den Bercken, & Essers, 1990).

The process errors in successive vectors $\mathbf{w}_t$ and the measurement errors in successive vectors $\mathbf{v}_t$ are assumed to have (a) zero expectations: $E(\mathbf{w}_t) = E(\mathbf{v}_t) = \mathbf{0}$ for all t, (b) zero covariances between vectors: $E(\mathbf{w}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all t and t', $E(\mathbf{w}_t\ \mathbf{w}'_{t'}) = E(\mathbf{v}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all $t \neq t'$ (nonzero variances and covariances for errors within vectors are in $\mathbf{Q}_t$ and $\mathbf{R}_t$), and (c) zero covariances with the initial state: $E(\mathbf{w}_t\ \mathbf{x}'_{t_0}) = E(\mathbf{v}_t\ \mathbf{x}'_{t_0}) = \mathbf{0}$ for all t. Further, (d) the error vectors and the initial state are assumed to be jointly multinormally distributed. Finally, it is assumed (e) $E(\mathbf{x}_{t_0}) = E(\mathbf{y}_{t_0}) = \mathbf{0}$, implying $E(\mathbf{x}_t) = E(\mathbf{y}_t) = \mathbf{0}$ for all t. (see Meditch, 1969, pp. 168-169). Assumptions (a) through (d) are essential in LISREL modeling as well as Kalman smoothing. Assumption (e) leads to the so-called 'zero means' LISREL model. In dropping this assumption, the 'structured means' LISREL model can be obtained, which enables the estimation of the mean structure in addition to the covariance structure (Jöreskog & Sörbom, 1989, p. 273). Accordingly, the SSM as well as the Kalman smoother are to be extended by the specification of additional input-effects (Lewis, 1986, pp. 69, 134; Oud et al., 1990, pp. 399-400; Oud et al., 1993). This paper, however, is restricted to the 'zero means' LISREL model and to the zero-input SSM.

Instead of using the full LISREL model, comprised of three equations and eight parameters matrices, the following submodel is used, comprised of only two equations and four parameter matrices (Jöreskog & Sörbom, 1989, p. 10).

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \qquad \text{with} \qquad cov(\boldsymbol{\zeta}) = \boldsymbol{\Psi}\ , \tag{2.3}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{with} \qquad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}\ . \tag{2.4}$$

Somewhat paradoxically, all conceivable LISREL models can be put in Equations 2.3 and 2.4 by combining all observed variables in $\mathbf{y}$ and all latent variables in $\boldsymbol{\eta}$. Equations 2.3 and 2.4, in fact, represent a more general model than the full LISREL model (Jöreskog & Sörbom, 1989, p. 190). By taking $\boldsymbol{\eta} = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1} \cdots \mathbf{x}'_{t_0+T-1}]'$ and $\mathbf{y} = [\mathbf{y}'_{t_0}\ \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_{t_0+T-1}]'$ with t_0 the initial time point and T the total number of time points considered, and putting the parameter matrices of Equations 2.1 and 2.2 on the appropriate places in the LISREL parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Theta}$, the LISREL model is easily formulated as a SSM. Notice, that the initial state $\mathbf{x}_{t_0}$, being exogenous or unexplained in the SSM, has its covariance matrix $\boldsymbol{\Phi}_{t_0} = E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0})$ specified in $\boldsymbol{\Psi}$. The other nonzero elements of $\boldsymbol{\Psi}$ are the process error variances and covariances in successive matrices $\mathbf{Q}_t$ with $t = t_0, \ldots, t_0 + T - 2$. Because all and only all the assumptions of the SSM are specified in the LISREL model, the LISREL model becomes fully equivalent to the SSM.

2.3 Maximum likelihood estimation of the LISREL model

Several estimation methods can be used in the LISREL program (Jöreskog & Sörbom, 1989, p. 16). Here we apply the ML method which maximizes the loglikelihood function of the free parameters in parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$, for given data in $\mathbf{Y}$:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{pN}{2}\log 2\pi \ . \tag{2.5}$$

$\boldsymbol{\theta}$ in Equation 2.5 contains the free parameters, $\mathbf{Y}_{p\times N}$ is the data matrix (N columns of independent replications of the p-variate vector $\mathbf{y}$, typically originating from a sample of randomly drawn subjects), $\boldsymbol{\Sigma}_{p\times p}$ is the model implied covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \ , \tag{2.6}$$

which is a function $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$, and $\mathbf{S}_{p\times p} = \frac{1}{N}\mathbf{Y}\ \mathbf{Y}'$ is the sample covariance matrix. The ML-estimator $\widehat{\boldsymbol{\theta}} = \mathrm{argmax}\ \ell(\boldsymbol{\theta}|\mathbf{Y})$ chooses that value of $\boldsymbol{\theta}$ which maximizes $\ell(\boldsymbol{\theta}|\mathbf{Y})$. However, instead of maximizing Equation 2.5 the LISREL program minimizes fit function

$$F_{ML} = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log|\mathbf{S}| - p \ , \tag{2.7}$$

with the same result. Equation 2.7 only differs from Equation 2.5 in the negative multiplying constant $-\frac{2}{N}$ and an additive constant; $\mathbf{S}$ is based on the data and thus constant in the LISREL fit function.

2.4 Kalman smoother

The Kalman smoother (Jazwinski, 1970; Lewis, 1986; Rauch, Tung, & Striebel, 1965) uses past, present, and future information for optimally estimating the latent state $\mathbf{x}_t$ (Equation 2.1) for an individual subject on the basis of its specific data sequence $\mathbf{y}_{t_0}, \mathbf{y}_{t_0+1}, \ldots, \mathbf{y}_{t_0+T-1}$. Writing $t_0 + T - 1 = s$, the optimal estimator or Kalman smoother turns out to be the conditional expectation (Meditch, 1969, p. 206)

$$\hat{\mathbf{x}}_t^s = E(\mathbf{x}_t|[\mathbf{y}_{t_0}'\ \mathbf{y}_{t_0+1}' \cdots \ \mathbf{y}_s']', \boldsymbol{\theta}) \qquad \text{with} \quad t \le s \ , \tag{2.8}$$

which depends on the data $[\mathbf{y}_{t_0}'\mathbf{y}_{t_0+1}' \cdots \mathbf{y}_s']'$ and via the model on the parameter values $\boldsymbol{\theta}$. It is obtained by minimizing the trace of the error covariance matrix (Shumway & Stoffer, 1982, p. 256)

$$\mathbf{P}_t^s = E(\mathbf{e}_t^s\ \mathbf{e}_t^{s\prime}|[\mathbf{y}_{t_0}'\ \mathbf{y}_{t_0+1}' \cdots \ \mathbf{y}_s']', \boldsymbol{\theta}) \ , \tag{2.9}$$

for estimation error $\mathbf{e}_t^s = \mathbf{x}_t - \hat{\mathbf{x}}_t^s$. The conditional covariance matrix $\mathbf{P}_t^s$ turns out to be independent of the data and can therefore be written as the unconditional covariance matrix: $\mathbf{P}_t^s = E(\mathbf{e}_t^s\ \mathbf{e}_t^{s\prime})$.

The Kalman smoother estimates can be calculated as follows. First, the discrete-time Kalman filter $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^t$ is run in a forward recursion. The Kalman filter differs from the Kalman smoother in neglecting future information for the estimation of $\mathbf{x}_t$, optimizing only over the information at time points $t' \leq t$. The Kalman filter in turn uses the predictor $\hat{\mathbf{x}}_{t-} = \hat{\mathbf{x}}_t^{t-1}$ (based on and optimal only for the information from past time points $t' \leq t-1$) with associated error covariance matrix $\mathbf{P}_{t-} = E(\mathbf{e}_{t-}\ \mathbf{e}_{t-}')$ for prediction error $\mathbf{e}_{t-} = \mathbf{x}_t - \hat{\mathbf{x}}_{t-}$. These are shown in Equations 2.10 and 2.11.

$$
\begin{aligned}
\hat{\mathbf{x}}_{t-} &= \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1}\ , &(2.10)\\
\mathbf{P}_{t-} &= \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}' + \mathbf{Q}_{t-1}\ . &(2.11)
\end{aligned}
$$

The Kalman filter measurement update $\hat{\mathbf{x}}_t$ and associated error covariance matrix $\mathbf{P}_t = E(\mathbf{e}_t\ \mathbf{e}_t')$ for filter error $\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$ are contained in Equations 2.14 and 2.15. Using the Kalman gain weighting matrix $\mathbf{H}_t$

$$
\mathbf{H}_t = \mathbf{P}_{t-}\mathbf{C}_t'(\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t)^{-1}\ , \tag{2.12}
$$

and defining $\mathbf{M}_t$:

$$
\mathbf{M}_t = \mathbf{I} - \mathbf{H}_t\mathbf{C}_t\ , \tag{2.13}
$$

the Kalman filter equations can be written as follows

$$
\begin{aligned}
\hat{\mathbf{x}}_t &= \mathbf{M}_t\hat{\mathbf{x}}_{t-} + \mathbf{H}_t\mathbf{y}_t\ , &(2.14)\\
\mathbf{P}_t &= \mathbf{M}_t\mathbf{P}_{t-}\mathbf{M}_t' + \mathbf{H}_t\mathbf{R}_t\mathbf{H}_t'\ . &(2.15)
\end{aligned}
$$

Also important in the missing data procedure is the Kalman filter error covariance matrix of errors $\mathbf{e}_t$ and $\mathbf{e}_{t-k}$ at different points in time:

$$
\begin{aligned}
\mathbf{P}_{t,t-k} &= \mathbf{M}_t\mathbf{P}_{t-,t-k}\\
&= \mathbf{M}_t\mathbf{A}_{t-1}\ \mathbf{M}_{t-1}\mathbf{A}_{t-2}\ \ldots\ \mathbf{M}_{t-k+1}\mathbf{A}_{t-k}\ \mathbf{P}_{t-k}\ . &(2.16)
\end{aligned}
$$

Its derivation is based on the recursive relation $\mathbf{e}_t = \mathbf{M}_t\mathbf{A}_{t-1}\mathbf{e}_{t-1} + \mathbf{M}_t\mathbf{w}_{t-1} - \mathbf{H}_t\mathbf{v}_t$.

At the initial time point $t = t_0$ where the first data are present and no past information is available, an estimator can be derived from Equation 2.14, assuming $\mathbf{A}_{t-1} = \mathbf{0}$. This implies $\hat{\mathbf{x}}_{t-} = E(\mathbf{x}_t)$, $\mathbf{P}_{t-} = \mathbf{\Phi}_t = cov(\mathbf{x}_t)$ which results in:

$$
\hat{\mathbf{x}}_t = \mathbf{M}_t E(\mathbf{x}_t) + \mathbf{H}_t\mathbf{y}_t\ . \tag{2.17}
$$

By substituting $\mathbf{\Phi}_t$ for $\mathbf{P}_{t-}$ in Equations 2.12 and 2.15, the weighting matrix $\mathbf{H}_t$ and the initial error covariance matrix $\mathbf{P}_t$ can be computed. Writing Equations 2.12 and 2.15 in another form,

$$\begin{aligned} \mathbf{H}_t &= \mathbf{P}_t\mathbf{C}_t'\mathbf{R}_t^{-1} \,, && (2.18)\\ \mathbf{P}_t &= (\mathbf{P}_{t-}^{-1} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1} \\ &= \mathbf{P}_{t-}(\mathbf{I} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t\mathbf{P}_{t-})^{-1} \,, && (2.19) \end{aligned}$$

it can be seen that Equations 2.17-2.19 for $E(\mathbf{x}_t) = \mathbf{0}$ (as assumed at initial time point $t = t_0$) and $\mathbf{P}_{t-} = \mathbf{\Phi}_t$, represent the well-known cross-sectional regression estimator (Lawley & Maxwell, 1971; Oud et al., 1993). This, and more generally Equations 2.17-2.19 for $E(\mathbf{x}_t) \neq \mathbf{0}$, is optimal for information at the single point in time t only (Singer, 1992, p. 141).

Second, the backward smoother algorithm is applied. It consists of three equations with $\mathbf{F}_t$ the backward gain matrix:

$$\begin{aligned} \hat{\mathbf{x}}_t^s &= \hat{\mathbf{x}}_t + \mathbf{F}_t(\hat{\mathbf{x}}_{t+1}^s - \hat{\mathbf{x}}_{(t+1)-}) \,, && (2.20)\\ \mathbf{P}_t^s &= \mathbf{P}_t + \mathbf{F}_t(\mathbf{P}_{t+1}^s - \mathbf{P}_{(t+1)-})\mathbf{F}_t' \,, && (2.21)\\ \mathbf{F}_t &= \mathbf{P}_t\mathbf{A}_t'(\mathbf{P}_{(t+1)-})^{-1} \,. && (2.22) \end{aligned}$$

The Kalman smoother state estimate $\hat{\mathbf{x}}_t^s$ (Equation 2.20) and Kalman smoother error covariance matrix $\mathbf{P}_t^s$ (Equation 2.21) require the Kalman filter estimate $\hat{\mathbf{x}}_t$ and covariance matrix $\mathbf{P}_t$ with $\hat{\mathbf{x}}_s$ and $\mathbf{P}_s$ as initial conditions (Jazwinski, 1970, p. 217; Lewis, 1986, p. 134; Rauch et al., 1965, p. 1447).

2.5 Inserting the Kalman smoother in the model implied covariance matrix

The model implied covariance matrix $\mathbf{\Sigma} = E(\mathbf{y}\ \mathbf{y}')$ of the LISREL model can be expressed in terms of the Kalman smoother $\hat{\mathbf{x}}_t^s$. The diagonal blocks of $\mathbf{\Sigma}$, obeying Equation 2.2, are

$$E(\mathbf{y}_t\ \mathbf{y}_t') = \mathbf{C}_t E(\mathbf{x}_t\ \mathbf{x}_t')\mathbf{C}_t' + \mathbf{R}_t \,. \qquad (2.23)$$

Rewriting $E(\mathbf{x}_t\ \mathbf{x}_t')$ as $E[(\hat{\mathbf{x}}_t^s + \mathbf{e}_t^s)(\hat{\mathbf{x}}_t^{s\prime} + \mathbf{e}_t^{s\prime})] = E(\hat{\mathbf{x}}_t^s\ \hat{\mathbf{x}}_t^{s\prime}) + \mathbf{P}_t^s$ in view of Kalman smoother property $E(\hat{\mathbf{x}}_t^s\ \mathbf{e}_t^{s\prime}) = \mathbf{0}$ (Jazwinski, 1970, p. 217 -218; Meditch, 1969, p. 209) one finds

$$E(\mathbf{y}_t\ \mathbf{y}_t') = \mathbf{C}_t E(\hat{\mathbf{x}}_t^s\ \hat{\mathbf{x}}_t^{s\prime})\mathbf{C}_t' + \mathbf{C}_t\mathbf{P}_t^s\mathbf{C}_t' + \mathbf{R}_t \,. \qquad (2.24)$$

For the off-diagonal blocks $E(\mathbf{y}_t\ \mathbf{y}_{t-k}')$ it is observed that $E(\mathbf{x}_t\ \mathbf{x}_{t-k}') = E[(\hat{\mathbf{x}}_t^s + \mathbf{e}_t^s)(\hat{\mathbf{x}}_{t-k}^{s\prime} + \mathbf{e}_{t-k}^{s\prime})] = E(\hat{\mathbf{x}}_t^s\ \hat{\mathbf{x}}_{t-k}^{s\prime}) + E(\mathbf{e}_t^s\ \mathbf{x}_{t-k}')$ because of Kalman smoother property

$E(\hat{\mathbf{x}}_t^s \, \mathbf{e}_{t-k}^{s\prime}) = \mathbf{0}$ (Jazwinski, 1970, p. 217), which leads to

$$E(\mathbf{y}_t \, \mathbf{y}_{t-k}') = \mathbf{C}_t E(\hat{\mathbf{x}}_t^s \, \hat{\mathbf{x}}_{t-k}^{s\prime})\mathbf{C}_{t-k}' + \mathbf{C}_t \mathbf{G}_{t,t-k}^s \mathbf{C}_{t-k}' \; . \tag{2.25}$$

Analogously to $\mathbf{P}_t^s$, $\mathbf{G}_{t,t-k}^s \equiv E(\mathbf{e}_t^s \, \mathbf{x}_{t-k}')$ may be determined in a backward recursion from $\mathbf{G}_{t+1,t-k}^s$

$$\mathbf{G}_{t,t-k}^s = \mathbf{F}_t \mathbf{G}_{t+1,t-k}^s + (\mathbf{I} - \mathbf{F}_t \mathbf{A}_t)\mathbf{P}_{t,t-k} \; , \tag{2.26}$$

starting with $\mathbf{G}_{t+1,t-k}^s = \mathbf{G}_{s,s-k-1}^s = E(\mathbf{e}_s \, \mathbf{x}_{s-k-1}') = E(\mathbf{e}_s \, \mathbf{e}_{s-k-1}') + E(\mathbf{e}_s \, \hat{\mathbf{x}}_{s-k-1}') = \mathbf{P}_{s,s-k-1}$ because of Kalman filter property $E(\mathbf{e}_s \, \hat{\mathbf{x}}_{s-k-1}') = \mathbf{0}$. Equation 2.26 results from the fact that the smoother error $\mathbf{e}_t^s$ follows the same backward recursion as the smoother estimate $\hat{\mathbf{x}}_t^s$

$$\mathbf{e}_t^s = \mathbf{e}_t + \mathbf{F}_t(\mathbf{e}_{t+1}^s - \mathbf{e}_{(t+1)-}) \; ,$$

and that

$$\mathbf{e}_{(t+1)-} = \mathbf{x}_{t+1} - \hat{\mathbf{x}}_{(t+1)-} = \mathbf{A}_t(\mathbf{x}_t - \hat{\mathbf{x}}_t) + \mathbf{w}_t = \mathbf{A}_t \mathbf{e}_t + \mathbf{w}_t \; ,$$

implying

$$E(\mathbf{e}_t^s \, \mathbf{x}_{t-k}') = E(\mathbf{e}_t \, \mathbf{x}_{t-k}') + \mathbf{F}_t[E(\mathbf{e}_{t+1}^s \, \mathbf{x}_{t-k}') - \mathbf{A}_t E(\mathbf{e}_t \, \mathbf{x}_{t-k}')] \; .$$

Equation 2.26 follows because of

$$E(\mathbf{e}_t \, \mathbf{x}_{t-k}') = \mathbf{P}_{t,t-k} \; .$$

Note that $\mathbf{G}_{t,t-k}^s = E(\mathbf{e}_t^s \, \mathbf{x}_{t-k}^{s\prime}) = E(\mathbf{e}_t^s \, \mathbf{e}_{t-k}^{s\prime}) = \mathbf{P}_{t,t-k}^s$ because of Kalman smoother property $E(\mathbf{e}_t^s \, \hat{\mathbf{x}}_{t-k}^{s\prime}) = \mathbf{0}$.

The relevance of Equations 2.24 and 2.25 for the missing data problem and the implementation of the EM algorithm is that they show the expectations of observed $\mathbf{y}_t \, \mathbf{y}_t'$ and $\mathbf{y}_t \, \mathbf{y}_{t-k}'$ to be equal to the expectations of $\mathbf{C}_t \hat{\mathbf{x}}_t^s \, \hat{\mathbf{x}}_t^{s\prime} \mathbf{C}_t{}' + \mathbf{C}_t \mathbf{P}_t^s \mathbf{C}_t{}' + \mathbf{R}_t$ and $\mathbf{C}_t \hat{\mathbf{x}}_t^s \, \hat{\mathbf{x}}_{t-k}^{s\prime} \mathbf{C}_{t-k}' + \mathbf{C}_t \mathbf{G}_{t,t-k}^s \mathbf{C}_{t-k}'$. The latter quantities are computable for individual subjects on the basis of complete as well as incomplete data. In the case of missing data at time points t' in the forward recursion, $\hat{\mathbf{x}}_{t'}$ and $\mathbf{P}_{t'}$ (Equations 2.14 and 2.15) are simply replaced by predictive values $\hat{\mathbf{x}}_{t'-}$ and $\mathbf{P}_{t'-}$ (Equations 2.10 and 2.11).

2.6 EM algorithm

The basic idea of the application of the EM algorithm is to estimate in the E-step the latent states and missing parts of the data by means of the Kalman smoother, based on the available data and parameter estimates from the M-step. Then, in the next M-step, parameters are estimated on the basis of the completed data from the E-step. These steps are iterated until the parameter estimates $\widehat{\boldsymbol{\theta}}$ converge.

Instead of the complete data loglikelihood function $\ell(\boldsymbol{\theta}|\mathbf{Y}) = \ell(\boldsymbol{\theta}|Y_{obs}, Y_{mis})$ in Equation 2.5, the loglikelihood function $\ell(\boldsymbol{\theta}|Y_{obs})$, given the observed data, must be maximized. As this cannot be done directly, the pseudo-likelihood or conditional loglikelihood expectation is determined. It depends on the observed data Y_{obs} and parameter values $\widehat{\boldsymbol{\theta}}_r$ of the preceding M-step. The EM algorithm thus consists of two steps:

E-step: If $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_r$ in the r^{th} iteration of the algorithm, determine the pseudo-likelihood

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_r) = E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r] - E_{Y_{mis}}[\ell(\widehat{\boldsymbol{\theta}}_r|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r] \ . \tag{2.27}$$

The expectation is taken over the distribution of the missing data Y_{mis} given the observed data Y_{obs} and the current estimate $\widehat{\boldsymbol{\theta}}_r$.

M-step:

$$\widehat{\boldsymbol{\theta}}_{r+1} = \text{argmax } Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_r) = \text{argmax } E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r] \ . \tag{2.28}$$

As the second term in the pseudo-likelihood (Equation 2.27) does not depend on $\boldsymbol{\theta}$, it suffices to maximize the conditional loglikelihood expectation $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r]$. It can be proven that iteration of these steps yields successive pairs of estimates $\widehat{\boldsymbol{\theta}}_r$ and $\widehat{\boldsymbol{\theta}}_{r+1}$ with non-decreasing loglikelihoods $\ell(\widehat{\boldsymbol{\theta}}_r|Y_{obs})$ and $\ell(\widehat{\boldsymbol{\theta}}_{r+1}|Y_{obs})$, converging to $\widehat{\boldsymbol{\theta}} = \text{argmax } \ell(\boldsymbol{\theta}|Y_{obs})$ (see Dembo & Zeitouni, 1986; Dempster et al., 1977; Little & Rubin, 1987; Singer, 1993).

For implementation of the EM algorithm the conditionally expected moment or covariance matrix $\mathbf{S}_{r+1} = E_{Y_{mis}}(\mathbf{S}|Y_{obs}, \widehat{\boldsymbol{\theta}}_r)$ is to be calculated in the E-step and inserted for $\mathbf{S}$ in Equation 2.5. This is due to the fact that the loglikelihood function in Equation 2.5 is linear in $\mathbf{S}$. Except for the replacement of $\mathbf{S}$ by $\mathbf{S}_{r+1}$, $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r]$ does not differ from $\ell(\boldsymbol{\theta}|\mathbf{Y})$ in Equation 2.5 (which is handled in the LISREL program by means of Equation 2.7, where again $\mathbf{S}$ is to be replaced by $\mathbf{S}_{r+1}$). Note, that in case of no missing data $E_{Y_{mis}}(\mathbf{S}|Y_{obs}, \widehat{\boldsymbol{\theta}}_r) = \mathbf{S}$, and $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \widehat{\boldsymbol{\theta}}_r]$ becomes equal to $\ell(\boldsymbol{\theta}|\mathbf{Y})$ of Equation 2.5.

Because missing data occur in subject specific patterns, dependent on the observed data vector $\mathbf{y}_{i,obs}$ for subject i, the conditional expectations $\hat{\mathbf{x}}_t^s$ and $\hat{\mathbf{x}}_{t-k}^s$ and covariance matrices $\mathbf{P}_t^s$ and $\mathbf{G}_{t,t-k}^s$ (see Equations 2.21 and 2.26) become subject specific too. They depend on the specific history of time points at which Kalman filter predictions (missing data: $\mathbf{H}_t = \mathbf{0} \Rightarrow \mathbf{M}_t = \mathbf{I}$) and Kalman filter measurement updates (observed data: $\mathbf{H}_t \neq \mathbf{0} \Rightarrow \mathbf{M}_t \neq \mathbf{I}$) are to be made. As $\mathbf{S}$

can be written as the mean of subject specific matrices $\mathbf{S}_i$:

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{S}_i = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\, \mathbf{y}_i' , \tag{2.29}$$

only the $\mathbf{S}_i$ of subjects with missing data, and only at the places where missing data occur, need to be changed to $\mathbf{S}_{i,r+1}$ for the calculation of $\mathbf{S}_{r+1}$. In each $\mathbf{S}_{i,r+1}$ the missing blocks $\mathbf{y}_{it}\, \mathbf{y}_{it}'$ and $\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}'$ are filled out with conditional expectations

$$E(\mathbf{y}_{it}\, \mathbf{y}_{it}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) = \mathbf{C}_t \hat{\mathbf{x}}_{it}^s\, \hat{\mathbf{x}}_{it}^{s\prime} \mathbf{C}_t' + \mathbf{C}_t \mathbf{P}_{it}^s \mathbf{C}_t' + \mathbf{R}_t , \tag{2.30}$$

and

$$E(\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) = \mathbf{C}_t \hat{\mathbf{x}}_{it}^s\, \hat{\mathbf{x}}_{i,t-k}^{s\prime} \mathbf{C}_{t-k}' + \mathbf{C}_t \mathbf{G}_{it,t-k}^s \mathbf{C}_{t-k}' . \tag{2.31}$$

For simplicity we omitted subscript r at the right hand sides of Equations 2.30-2.31. Blocks combining observed $\mathbf{y}_t$ with missing $\mathbf{y}_{t-k}$ or missing $\mathbf{y}_t$ with observed $\mathbf{y}_{t-k}$ become, respectively, for subject i:

$$\begin{aligned} E(\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) &= \mathbf{y}_{it}\, E(\mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) \\ &= \mathbf{y}_{it}\, \hat{\mathbf{x}}_{i,t-k}^{s\prime} \mathbf{C}_{t-k}' , \end{aligned} \tag{2.32}$$

and

$$\begin{aligned} E(\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}\, , \widehat{\boldsymbol{\theta}}_r) &= E(\mathbf{y}_{it} | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r)\, \mathbf{y}_{i,t-k}' \\ &= \mathbf{C}_t \hat{\mathbf{x}}_{it}^s\, \mathbf{y}_{i,t-k}' . \end{aligned} \tag{2.33}$$

Above we tacitly assumed that the observations within vectors $\mathbf{y}_t$ are either completely missing or not missing at all. However, it can happen that the observations at t, $t-k$ or both are only partially missing. As a consequence, only part of the information can be used in the Kalman filter measurement update (Equation 2.14) or initial estimator (Equation 2.17). This is done by entering zeroes in these equations for the partially missing observations in $\mathbf{y}_t$, and fixing the corresponding factorloading rows of the $\mathbf{C}_t$ matrix at zero (Shumway & Stoffer, p. 259) in the computation of matrices $\mathbf{H}_t$ (Equation 2.12) and $\mathbf{M}_t$ (Equation 2.13). In this way only the available measurement information is processed by the Kalman filter and smoother. If $\mathbf{R}_{(12)t} = \mathbf{0}$ (no measurement error correlations between observed part $\mathbf{y}_{(1)t}$ and missing part $\mathbf{y}_{(2)t}$ in each $\mathbf{y}_t$). Equations 2.30-2.33 remain valid (with the full matrices $\mathbf{C}_t$ and $\mathbf{C}_{t-k}$ entered), except of course that the nonmissing elements in the blocks $\mathbf{y}_t\, \mathbf{y}_t'$ and $\mathbf{y}_t\, \mathbf{y}_{t-k}'$ are not to be replaced by their conditional expectations and that the elements combining missing with nonmissing data are handled analogously to Equations 2.32-2.33. $\mathbf{R}_{(12)t} \neq \mathbf{0}$ requires a modification of Equation 2.30, which is given by Shumway and Stoffer (1982, p. 257).

An interesting alternative for the maximization of $\ell(\boldsymbol{\theta}|\mathbf{Y})$ in Equation 2.5 is offered by Singer (1993). Employing EM even when $\mathbf{Y}$ contains complete data, it allows to replace the LISREL program by any observed variables SEM program that uses the maximum likelihood method. Combining all latent states in the $mT \times N$ matrix $\mathbf{X}$ (m the number of state variables in state vector $\mathbf{x}_t$ and mT the total number of state variables in $\boldsymbol{\eta}$), it consists of considering the latent unknown $\mathbf{X}$ as the missing data. Bayes' formula enables one to decompose the complete data loglikelihood $\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})$ into separate loglikelihoods for the structural (state equation) and measurement (output equation) parts of the LISREL model:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X}) &= \ell(\boldsymbol{\theta}|\mathbf{X}) + \ell(\boldsymbol{\theta}|\mathbf{Y}|\mathbf{X}) \\ &= \ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) + \ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) \ . \end{aligned} \tag{2.34}$$

Writing $\boldsymbol{\Sigma}_\eta = (\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \boldsymbol{B}')^{-1}$, $\mathbf{S}_\eta = 1/N\mathbf{X}\ \mathbf{X}'$, and $\mathbf{S}_\varepsilon = 1/N(\mathbf{Y} - \boldsymbol{\Lambda}\mathbf{X})(\mathbf{Y}-\boldsymbol{\Lambda}\mathbf{X})' = \mathbf{S}+\boldsymbol{\Lambda}\mathbf{S}_\eta\boldsymbol{\Lambda}'-(1/N)\mathbf{Y}\mathbf{X}'\boldsymbol{\Lambda}'-(1/N)\boldsymbol{\Lambda}\mathbf{X}\mathbf{Y}'$ one derives the separate loglikelihoods:

$$\begin{aligned} \ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) &= -\frac{N}{2}\log|\,\boldsymbol{\Sigma}_\eta\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}_\eta\boldsymbol{\Sigma}_\eta^{-1}) - \frac{mTN}{2}\log 2\pi \ , \\ \ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) &= -\frac{N}{2}\log|\,\boldsymbol{\Theta}\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}_\varepsilon\boldsymbol{\Theta}^{-1}) - \frac{pN}{2}\log 2\pi \ . \end{aligned} \tag{2.35}$$

The measurement part in Equation 2.36 now takes the form of a regression analysis problem.

The conditional loglikelihood expectation to be maximized iteratively in EM decomposes accordingly:

$$\begin{aligned} E_X[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r] &= E_X[\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X})|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r] + \\ &\quad\ E_X[\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X})|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r] \ . \end{aligned} \tag{2.36}$$

Implementation of the EM algorithm by maximizing Equation 2.36 in the M-step requires substitution of the unknown $\mathbf{S}_\eta$ and $\mathbf{X}$ in Equations 2.35 and 2.36 by $\mathbf{S}_{\eta,r+1} = E_X(\mathbf{S}_\eta|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r)$ and $\mathbf{X}_{r+1} = E_X(\mathbf{X}|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r)$, and leads to the Kalman smoother in the E-step. $\mathbf{S}_{\eta,r+1}$ consists of diagonal blocks

$$\frac{1}{N}\sum_{i=1}^{N}[\hat{\mathbf{x}}_{it}^s\ \hat{\mathbf{x}}_{it}^{s\prime} + \mathbf{P}_{it}^s] \ ,$$

and off-diagonal blocks

$$\frac{1}{N}\sum_{i=1}^{N}[\hat{\mathbf{x}}_{it}^s\ \hat{\mathbf{x}}_{i,t-k}^{s\prime} + \mathbf{G}_{it,t-k}^s] \ . \tag{2.37}$$

Again it can be proven that (Singer, 1993) the successive EM estimates $\widehat{\boldsymbol{\theta}}_{r+1} = [\widehat{\boldsymbol{\theta}}'_{B,\Psi}\ \widehat{\boldsymbol{\theta}}'_{\Lambda,\Theta}]'_{r+1} = \text{argmax}\ E_X[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r]$ converge to the maximum likelihood estimate argmax $\ell(\boldsymbol{\theta}|\mathbf{Y})$. This implies that the separate maximization of Equations 2.35 and 2.36 (e.g. by means of the LISREL program but also by means of any observed variables SEM program), using and iteratively inserting Kalman smoother estimates, leads to the same parameter estimates as the direct maximization of Equation 2.5 by means of the LISREL program. Disadvantages of the indirect decomposition procedure are that no cross-restrictions between the parameters in $\boldsymbol{\theta}_{B,\Psi}$ and $\boldsymbol{\theta}_{\Lambda,\Theta}$ are possible and that the Kalman smoother estimates need to be computed for all subjects, not only for subjects with missing data in $\mathbf{Y}$. In the case of missing data in $\mathbf{Y}$, $E_X[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\widehat{\boldsymbol{\theta}}_r]$ in Equation 2.36 is replaced by $E_{Y_{mis},X}[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|Y_{obs},\widehat{\boldsymbol{\theta}}_r]$, $\mathbf{S}$ in Equation 2.36 by the matrix $\mathbf{S}_{r+1}$ explained previously, and blocks with missing $\mathbf{y}_{it}$ in $\mathbf{YX}'$ by $\mathbf{C}_t\hat{\mathbf{x}}^s_{it}\ \hat{\mathbf{x}}^{s\prime}_{it} + \mathbf{C}_t\mathbf{P}^s_{it}$, $\mathbf{C}_t\hat{\mathbf{x}}^s_{it}\ \hat{\mathbf{x}}^{s\prime}_{i,t-k} + \mathbf{C}_t\mathbf{G}^s_{it,t-k}$ and $\mathbf{C}_t\hat{\mathbf{x}}^s_{it}\ \hat{\mathbf{x}}^{s\prime}_{i,t+k} + \mathbf{C}_t\mathbf{G}^s_{it,t+k}$.

2.7 An educational research example

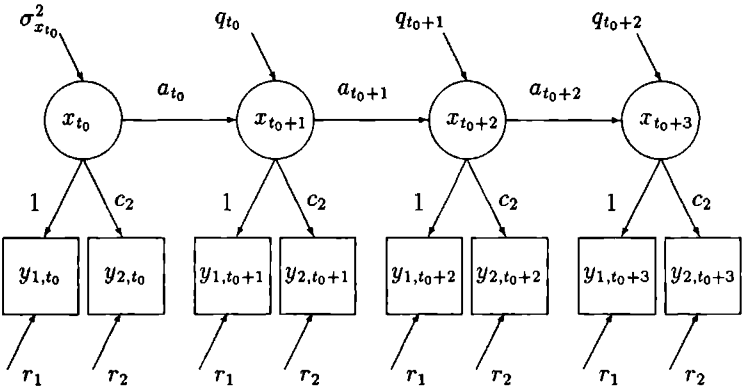
The missing data procedure was applied to a sample of 838 pupils from primary school for whom measurements with respect to the ability of decoding speed had been taken at four points in time, covering a two year period. At each point in time the two test forms A and B for measuring decoding speed (Brus & Voeten, 1979) had been used. The data set of eight observed variables contained 3.9% missing scores, only 740 or 88.3% of the 838 cases had complete observations.

The LISREL model, using state space notation, is shown in Figure 2.1. The squares represent the observed variables (tests A and B) at different points in time, loading on the four latent decoding speed variables represented by the circles. The factorloadings as well as the measurement error variances of the same test at different occasions were assumed to be equal. The latent variables were scaled by fixing the factorloadings of test A at one. Ten parameters had to be estimated, which resulted in a LISREL model with 26 degrees of freedom.

The EM algorithm was initialized by means of the LISREL solution of the N=740 complete cases analysis (CC). The results of both the EM and CC analysis are shown in Table 2.1. The EM algorithm took four iterations before convergence was reached. By convergence is meant that additional iterations did not result in any change of any of the LISREL parameter estimates.

The autoregressive parameter estimates of both the EM and CC solutions are very high, indicating that the latent decoding speed curves of individual pupils are almost parallel and regression toward the population mean negligible. For the EM solution, autoregression turns out to be somewhat higher between the first and second, and the second and third time point than for the CC solution. From the third to the fourth time point, CC shows the highest autoregression. Other

Fig. 2.1: LISREL model in state space form for decoding speed development of pupils.



differences between both solutions include the measurement error variances of the instruments being higher in EM than in CC.

The computation of the standard errors (indicated between parentheses in Table 2.1) is somewhat problematic in EM. In the last M-step of the algorithm, LISREL yields standard errors based on the ML covariance matrix estimate. The problem is, what value should be inserted for N in the denominator of the standard errors. The standard errors will become too small, if the total N (in the example N=838) is inserted. Little & Rubin (1987, p. 157) suggest using the N of the number of subjects which have observed values on all variables involved in the analysis, that is the CC N (in the example N=740). This will give an upper bound to the correct standard errors. Both sets of standard errors were computed, next to a set based on a compromise N, obtained by dividing the total number of nonmissing scores by the total number of observed variables in the model (in the example N=805). It results in approximate standard errors between the lower and upper bound.

It can be seen in Table 2.1 that the differences between the sets of standard errors are negligible and in several cases even not noticeable with the number of digits reported. Considering the large samples required in LISREL analyses, this will rather generally be the case and the use of different sets will seldom lead to different decisions. For safety, however, it is recommendable to compute all three sets and to check whether they indeed lead to the same decisions.

In order to illustrate the behavior of EM missing data procedure under different missing data mechanisms and amounts of missing data, a few simulations were conducted starting from the sample of complete cases: N = 740. Four missing data mechanisms (A, B, C, D) were applied and made to yield, respectively, 20% and 30% missing scores. The aim was to find out whether the EM algorithm was able to reproduce the known ML estimates of the N=740 CC analysis under each of these mechanism and percentage of missings conditions.

Tab. 2.1: LISREL parameter estimates for the model in Figure 2.1 from the EM and CC solution with standard errors between parentheses.

Latent autoregressive parameters	N		a_{t_0}	a_{t_0+1}	a_{t_0+2}
EM			0.952	0.959	0.916
	(838)		(0.015)	(0.014)	(0.012)
	(805)		(0.015)	(0.014)	(0.012)
	(740)		(0.016)	(0.015)	(0.012)
CC			0.948	0.952	0.921
	(740)		(0.016)	(0.015)	(0.013)
Latent initial variance and unexplained variances	N	$\sigma^2_{x_{t_0}}$	q_{t_0}	q_{t_0+1}	q_{t_0+2}
EM		188.74	24.69	20.64	12.70
	(838)	(9.55)	(1.66)	(1.43)	(1.07)
	(805)	(9.75)	(1.70)	(1.46)	(1.09)
	(740)	(10.17)	(1.77)	(1.52)	(1.13)
CC		180.68	24.71	20.89	13.38
	(740)	(9.72)	(1.74)	(1.50)	(1.14)
Factor loadings	N	c_1	c_2		
EM		1.000	1.017		
	(838)		(0.006)		
	(805)		(0.006)		
	(740)		(0.006)		
CC		1.000	1.024		
	(740)		(0.006)		
Measurement error variances	N	r_1	r_2		
EM		11.14	9.70		
	(838)	(0.46)	(0.45)		
	(805)	(0.47)	(0.46)		
	(740)	(0.49)	(0.48)		
CC		10.34	9.19		
	(740)	(0.46)	(0.46)		

In mechanism A missing scores were assigned to the cells of the data matrix on the basis of a simple random scheme. This was also done for mechanism B, but if a missing occurred at a specific point in time for a specific variable, all following scores were assigned missing too. As a result, some cases, which happened to be assigned a missing in the first variable at the first point in time, dropped from the data set. Both mechanisms A and B are examples of the MCAR (Missing Completely At Random) mechanism defined by Rubin (1976) and Little and Rubin (1987).

Mechanism C was applied for simulating a MAR (Missing At Random) mechanism. The key issue in the distinction between MCAR and MAR mechanisms is whether and how missingness is related to data values (Little, 1992, pp. 1229-1230). In contrast to a MCAR mechanism, in which missingness does not depend on data values at all, neither in Y_{mis} nor Y_{obs}, missingness in a MAR mechanism does depend on the data values in $\mathbf{Y}$ but only through the observed values in Y_{obs}. In the C mechanism missings were not directly assigned, but for each observed score below some criterion at time point t missings were assigned to both observed variables at time point $t+1$. The criterion used was the z-score for the variable at time point t in the sample of N=740 pupils. To obtain the desired percentages of 20% and 30% missing data z-scores of -0.13 and 0.66 respectively had to be used as criteria. In order to prevent the procedure to assign missings mainly at the second and fourth time point (caused by the fact that no missings are possible for mechanism C at the first time point), checking z-scores started randomly at the first or second time point.

Tab. 2.2: N of pupils having all data missing, N of initializing CC analysis, and number of EM iterations needed for convergence for each combination of missing data mechanism and percentage.

	N all missing	N initializing CC	No. of iterations
A 20% missing	0	131	10
B 20% missing	37	490	5
C 20% missing	0	304	11
D 20% missing	60	484	6
A 30% missing	0	54	14
B 30% missing	52	373	10
C 30% missing	0	117	24
D 30% missing	120	397	6

Finally, a non-MAR mechanism was applied to the data set. In a non-MAR mechanism missingness is dependent on the values of the very missing data. In non-MAR mechanism D every score below the z criterion value was assigned missing itself. z-scores of -0.90 and -0.64 had to be used for obtaining, respectively, 20% and 30% missing data. Because the variables are highly correlated, mechanism D also led to a number of pupils having all data missing (see the first column in Table 2.2).

The second column in Table 2.2 gives the numbers of complete cases left after application of the four mechanisms with each of both percentages of missings. The CC LISREL solutions based on those data sets were used to initialize the EM algorithm. Whereas MCAR and MAR mechanisms are ignorable in EM, non-MAR mechanisms are not. For this reason the EM results under mechanisms A, B, and C were expected to be satisfactory in contrast to those under mechanism D. The numbers of iterations needed for reaching convergence are given in the third column of Table 2.2. These are highest for mechanism C, under which EM had the formidable but, according to the theory, feasible task of adjusting the strongly deviating initial solution.

Tables 2.3 and 2.4 contain the results. As a measure of deviance of the parameter estimates, the 95% and 99% confidence intervals of the N=740 CC solution were used. All parameter estimates of the MCAR mechanisms A and B turned out to be within the 95% confidence intervals. For the C mechanism all parameter estimates were within the 99% confidence intervals, except for q_{t_0+1} in the 30% missing data case. In general, the results illustrate the expected behavior of missing data procedure. They are quite satisfactory for the MCAR and MAR mechanisms even with 30% missings. For the non-MAR mechanism D considerable deviations from the $N = 740$ CC solution resulted. Especially the autoregressive parameters and unexplained variances are strongly underestimated.

2.8 Discussion

The missing data procedure presented has several advantages compared to the so-called 'ad-hoc' methods (see Little & Rubin, chap. 4, 1987): (a) the statistically correct ML estimate is obtained, (b) each E-step of the EM algorithm produces a Gramian covariance matrix (in contrast to e.g. the 'covariance matrix ' based on so-called pairwise deletion), thus avoiding various kinds of improper solutions (Bentler & Jamshidian, 1994), and (c) all available measurement information is processed by the algorithm. In addition, it provides optimal values to be imputed for the missing values. If an appropriate model has been used, the resulting conditional expectations $\mathbf{C}_t\hat{\mathbf{x}}_{it}^s$ from the last E-step of the algorithm may be imputed. A stochastic extension (Little & Rubin, 1987, p. 61) consists in adding error components to the imputed values, drawn from a multinormal distribution with covariance matrix $\mathbf{C}_t\mathbf{P}_t^s\mathbf{C}_t' + \mathbf{R}_t$ (see Equation 2.30).

It has been observed that the EM algorithm is slow near the solution (Shumway & Stoffer, 1982, p. 255; Singer, 1990, 1992). The computing time needed for the LISREL jobs, however, can be shortened if in each M-step the parameter estimates of the preceding M-step are used as starting values. This is especially useful for complex models in which the computation of the initial estimates provided by the LISREL program demands a considerable amount of computing time.

Tab. 2.3: EM state equation parameter estimates for the model in Figure 2.1 under four missing data mechanisms and two missing data percentages, compared to the N=740 CC estimates.

Latent autoregressive parameters		a_{t_0}	a_{t_0+1}	a_{t_0+2}
CC		**0.948**	**0.952**	**0.921**
		(0.016)	**(0.015)**	**(0.013)**
A 20% missing		0.957	0.947	0.921
B 20% missing		0.955	0.947	0.917
C 20% missing		0.943	0.949	0.909
D 20% missing		0.888**	0.889**	0.898
A 30% missing		0.943	0.949	0.923
B 30% missing		0.951	0.951	0.906
C 30% missing		0.935	0.954	0.898
D 30% missing		0.880**	0.891**	0.901
Latent initial variance and unexplained variances	$\sigma^2_{x_{t_0}}$	q_{t_0}	q_{t_0+1}	q_{t_0+2}
CC	**180.68**	**24.71**	**20.89**	**13.38**
	(9.72)	**(1.74)**	**(1.50)**	**(1.14)**
A 20% missing	179.52	23.59	21.46	13.27
B 20% missing	180.06	22.75	21.32	13.50
C 20% missing	181.44	25 58	21.71	15.86*
D 20% missing	151.49**	19.87**	17.81*	11.66
A 30% missing	181.28	24.51	20.56	14.23
B 30% missing	180.49	25.24	21.37	13.18
C 30% missing	181.48	25.57	24.81**	15.84*
D 30% missing	146.56**	18.94**	17.01**	10.78*

* *Outside of 95% confidence interval around CC (N=740) estimate.*

** *Outside of 99% confidence interval around CC (N=740) estimate*

Tab. 2.4: EM state equation parameter estimates for the model in Figure 2.1 under four missing data mechanisms and two missing data percentages, compared to the N=740 CC estimates.

Factor loadings	c_1	c_2
CC	**1.000**	**1.024**
		(0.006)
A 20% missing		1.022
B 20% missing		1.019
C 20% missing		1.018
D 20% missing		1.020
A 30% missing		1.026
B 30% missing		1.019
C 30% missing		1.019
D 30% missing		1.017
Measurement error variances	r_1	r_2
CC	**10.34**	**9.19**
	(0.46)	**0.46)**
A 20% missing	10.37	9.57
B 20% missing	10.08	9.49
C 20% missing	10.52	9.78
D 20% missing	10.72	9.59
A 30% missing	9.90	9.56
B 30% missing	10.14	9.09
C 30% missing	9.87	10.26*
D 30% missing	10.85	9.85

* *Outside of 95% confidence interval around CC (N=740) estimate.*

** *Outside of 99% confidence interval around CC (N=740) estimate.*

Chapter 3

Handling missing data in the construction of the pupil monitoring system LISKAL[1]

Abstract

In an early version of the pupil monitoring system LISKAL, the Kalman filter was implemented for the estimation of individual latent developmental curves based on a longitudinal zero means SEM model. A pupil's developmental level was estimated relatively in terms of its deviation from the population zero mean development. In a recent improved LISKAL version individual latent developmental curves can be obtained on an absolute scale. The Kalman filter estimates are based on a structured means SEM model. In the case of missing data, the SEM model can be estimated using the EM (Expectation-Maximization) algorithm in conjunction with the Kalman smoother. It is shown which adjustments are necessary to implement this algorithm for the structured means SEM model. An example illustrates how the algorithm can be made useful in the construction of the pupil monitoring system LISKAL.

3.1 Introduction

The last decade, monitoring systems are increasingly used for assessing and diagnosing the development of pupils' abilities throughout the primary school period (Moelands, Mommers & Oud, 1990). In the pupil monitoring system LISKAL[2],

---

[1] Adapted version of Jansen, R A R G , & Oud, J H L (1995) Handling missing data in the construction of the pupil monitoring system LISKAL In I Partchev (Ed), *Proceedings of the SMABS 1994 conference Multivariate analysis in the behavioral sciences Philosophic to technical* (pp 39-48) Sofia Prof Marin Drinov Publishing House

[2] LISKAL refers to the combination of the LISREL program and the Kalman filter Although LISREL was employed in the construction of the pupil monitoring system LISKAL, other structural equation modeling or SEM programs may be used also

the Kalman filter is implemented for the estimation of individual latent developmental curves based on the maximum likelihood (ML) solution of a longitudinal SEM model (Oud, van den Bercken & Essers, 1990; Oud, van Leeuwe & Jansen, 1993). In an early version these estimates were based on the standardized solution of the so-called zero means SEM model (Oud, Mommers, Smitshuis, Doppenberg, Devilee & Heijmans, 1993). Intra-individual change could only be interpreted relatively in terms of deviations from the population's unknown mean development. In the new LISKAL version, individual latent developmental curves can be estimated on an absolute scale. The position of a pupil's level of achievement at a certain time point can be compared to previous or later time points in terms of growth or decay on an absolute scale. At the same time, its position can also be assessed relatively in comparison to the absolute latent mean growth curve. Hence, in Kalman filtering on the basis of a structured means SEM model (Jöreskog & Sörbom, 1989), the absolute as well as the relative position of a pupil's level of development can be assessed.

As has been noticed by many authors (Dempster, Laird & Rubin, 1977; Jöreskog, 1993; Little & Rubin, 1987; Rubin, 1976), missing data pose severe problems in obtaining correct ML estimates. These problems are solvable for a wide range of models by means of the EM algorithm. For the estimation of longitudinal SEM models from incomplete data, Jansen and Oud (1995) implemented the EM algorithm, combining a SEM program with the Kalman smoother. It is based on the zero means SEM model. The purpose of the paper is to show how the missing data procedure is implemented for the construction of the new LISKAL version, based on the structured means SEM model.

First, the structured means SEM model is formulated in terms of the SSM in order to make the Kalman smoother accessible for the proposed missing data procedure. Second, we summarize the SEM ML estimation method. Third, we present the modifications relevant for the missing data procedure in case Kalman smoothing is to be based on a structured means SEM model. For further details on the implementation of the EM algorithm we refer to Jansen and Oud (1995), Oud and Jansen (1996), Shumway and Stoffer (1982) and Singer (1990, 1992). Finally, the modified missing data procedure is applied to an educational research example to show how it can be employed for the construction of LISKAL. We conclude with a short comment on the results.

3.2 State space modeling and SEM

The linear stochastic discrete-time SSM consists of two parts: the dynamic part or state equation (Equation 3.1), and the static part or output equation (Equation 3.2),

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1} \quad \text{with} \quad cov(\mathbf{w}_{t-1}) = \mathbf{Q}_{t-1}\ , \tag{3.1}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \qquad \text{with} \qquad cov(\mathbf{v}_t) = \mathbf{R}_t \ . \tag{3.2}$$

The state equation describes the dependence of the latent state variables in the m-vector $\mathbf{x}_t$ on their lagged values in $\mathbf{x}_{t-1}$, with $\mathbf{A}_{t-1}$ $(m \times m)$ the system matrix, containing the autoregressive and cross-lagged effects between the state variables at successive discrete time points t and $t-1$: $t, t-1 \in \{t_0, t_0+1, \dots, t_0+T-1\}$ for integers t_0 and $T \geq 2$, with t_0 the initial time point and T the total number of time points considered. $\mathbf{B}_{t-1}$ $(m \times q)$ is the input-effect matrix with $\mathbf{u}_{t-1}$ $(q \times 1)$ representing the fixed input-variables. $\mathbf{w}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{t-1})$ is the m-vector of random forcing terms.

The instantaneous output equation connects the latent state variables to the observables in the p-vector $\mathbf{y}_t$. It is equivalent to the factor model equation in factor analysis with $\mathbf{C}_t$ $(p \times m)$ the factor pattern matrix. $\mathbf{D}_t$ $(p \times q)$ describes the influence of the fixed input-variables in the q-vector $\mathbf{u}_t$. Finally, $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ is a p-vector of random measurement errors.

For the error vectors, it is assumed that $E(\mathbf{w}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all t and t', $E(\mathbf{w}_t\ \mathbf{w}'_{t'}) = E(\mathbf{v}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all $t \neq t'$. Furthermore, $E(\mathbf{w}_t\ \mathbf{x}'_{t_0}) = E(\mathbf{v}_t\ \mathbf{x}'_{t_0}) = \mathbf{0}$ for all t with $\mathbf{x}_{t_0}$ the initial state vector. Finally, the error vectors and the initial state are jointly multinormally distributed.

While the model of Equations 3.1 and 3.2 represents the general case as regards the fixed input-variables in $\mathbf{u}_t$, only the special case of $u_t = 1$ for all t, being the so-called unit input-variable, is needed to define the structured means SEM model. As the unit input-variable is constant over times and common to all subjects in the sample, $\mathbf{B}_{t-1}$ and $\mathbf{D}_t$ become, respectively, $\mathbf{b}_{t-1}$ representing the latent growth intercepts, and $\mathbf{d}_t$ representing the location parameters (origins) of the measurement instruments. In another special case the input-variables are all constant over time ($\mathbf{u}_t = \mathbf{u}_{t-k}$ for all t and $k > 0$) but, apart from the unit input-variable, varying over subjects (e.g. gender). Finally, in the general case, additional input-variables are specified that vary over time points as well as over subjects.

For representing the SSM as a SEM model, a submodel of two equations and four parameter matrices is used instead of the full SEM model comprised of three equations and eight parameters matrices (Jöreskog & Sörbom, 1989).

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \qquad \text{with} \qquad cov(\boldsymbol{\zeta}) = \boldsymbol{\Psi}\ , \tag{3.3}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{with} \qquad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}\ . \tag{3.4}$$

Paradoxically, Equations 3.3 and 3.4 represent a more general model than the full SEM model. All conceivable SEM models can, in fact, be put in Equations 3.3 and 3.4 (Jöreskog & Sörbom, 1989, p. 10, 190). For deriving the SEM model and the form of its parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$, the SSM (Equations 3.1 and

3.2) is reformulated as follows,

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{t-1} & \mathbf{B}_{t-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{u}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{t-1} \\ \mathbf{u}_t \end{bmatrix}$$

$$\text{with} \quad E\left(\begin{bmatrix} \mathbf{w}_{t-1} \\ \mathbf{u}_t \end{bmatrix} [\mathbf{w}'_{t-1} \; \mathbf{u}'_t]\right) = \begin{bmatrix} \mathbf{Q}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{u}_t} \end{bmatrix}, \tag{3.5}$$

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{C}_t & \mathbf{D}_t \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} + \begin{bmatrix} \mathbf{v}_t \\ \mathbf{0} \end{bmatrix}$$

$$\text{with} \quad E\left(\begin{bmatrix} \mathbf{v}_t \\ \mathbf{0} \end{bmatrix} [\mathbf{v}'_t \; \mathbf{0}']\right) = \begin{bmatrix} \mathbf{R}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.6}$$

Putting the fixed input-variables in $\mathbf{u}$, possibly consisting of the unit input-variable only, and defining,

$$\begin{array}{lll}
\boldsymbol{\eta} = [\mathbf{x}' \; \mathbf{u}']' & \text{with} & \mathbf{x} \;= [\mathbf{x}'_{t_0} \; \mathbf{x}'_{t_0+1} \cdots \mathbf{x}'_{t_0+T-1}]', \\
\mathbf{y} = [\mathbf{y}'_\circ \; \mathbf{u}']' & \text{with} & \mathbf{y}_\circ = [\mathbf{y}'_{t_0} \; \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_{t_0+T-1}]', \\
\boldsymbol{\zeta} = [\mathbf{w}' \; \mathbf{u}']' & \text{with} & \mathbf{w} = [(\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0}))' \; \mathbf{w}'_{t_0} \cdots \mathbf{w}'_{t_0+T-2}]', \\
\boldsymbol{\varepsilon} = [\mathbf{v}' \; \mathbf{0}']' & \text{with} & \mathbf{v} \;= [\mathbf{v}'_{t_0} \; \mathbf{v}'_{t_0+1} \cdots \mathbf{v}'_{t_0+T-1}]',
\end{array}$$

the SEM model is derived. Notice, that the initial state $\mathbf{x}_{t_0}$, being exogenous or unexplained in the SSM, has its covariance matrix $\mathbf{\Phi}_{t_0}$ specified in $\mathbf{\Psi}$, and its mean at the place of $\boldsymbol{B}$ corresponding to the unit input-variable. The other nonzero elements of $\mathbf{\Psi}$ are the process error variances and covariances in successive matrices $\mathbf{Q}_t$ with $t = t_0, \ldots, t_0 + T - 2$, and $\mathbf{\Phi}_{\mathbf{u}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{u}_i \, \mathbf{u}'_i$.

3.3 Maximum likelihood estimation of the SEM model (M-step)

The ML estimation method is employed in the M-step of the EM algorithm (Jansen & Oud, 1995). For complete data, ML maximizes the loglikelihood function:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{N}{2}\log|\,\mathbf{\Sigma}\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \frac{pN}{2}\log 2\pi\,, \tag{3.7}$$

where $\boldsymbol{\theta}$ contains the free parameters in matrices $\boldsymbol{B}$, $\mathbf{\Psi}$, $\mathbf{\Lambda}$, and $\mathbf{\Theta}$. $\mathbf{Y}_{p\times N}$ is the data matrix (N replications of the p-variate vector $\mathbf{y}$), $\mathbf{\Sigma}_{p\times p} = E(\mathbf{y}\;\mathbf{y}')$ is the model implied augmented moment matrix expressed as follows

$$\mathbf{\Sigma} = \mathbf{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\mathbf{\Psi}(\mathbf{I} - \boldsymbol{B}')^{-1}\mathbf{\Lambda}' + \mathbf{\Theta}\,. \tag{3.8}$$

It is a function $\Sigma(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$. $\mathbf{S}_{p\times p} = \frac{1}{N}\mathbf{Y}\ \mathbf{Y}'$ is the sample augmented moment matrix. The ML-estimator is $\widehat{\boldsymbol{\theta}} = \text{argmax}\ \ell(\boldsymbol{\theta}|\mathbf{Y})$, choosing the value of $\boldsymbol{\theta}$ that maximizes $\ell(\boldsymbol{\theta}|\mathbf{Y})$. Instead of maximizing Equation 3.7, however, the SEM program minimizes fit function

$$F_{ML} = \log \mid \Sigma \mid + \text{tr}(\mathbf{S}\Sigma^{-1}) - \log \mid \mathbf{S} \mid - p\ , \tag{3.9}$$

with the same result. Equation 3.9 only differs from Equation 3.7 in the negative multiplying constant $-\frac{2}{N}$ and an additive constant; $\mathbf{S}$ is based on the data and thus invariant in the fit function. With q input-variables, the number of random variables in $\mathbf{y}$ is, in fact, $p-q$. However, whether $\mathbf{u}$ is considered random or fixed in a SEM analysis, makes no difference. In both cases, the input part $\boldsymbol{\Phi}_{\mathbf{u}}$ in Σ is set equal to the input part of $\mathbf{S}$ (Jöreskog & Sörbom, 1989, p. 140). That is, the input part $\boldsymbol{\Phi}_{\mathbf{u}}$ is just the mean product over the set of input values in the sample: $\frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i\ \mathbf{u}_i'$, and minimizing F_{ML} gives the maximum likelihood estimate.

3.4 Kalman smoothing based on the structured means SEM model (E-step)

The Kalman smoother uses past, present and future information for optimally estimating the latent states $\mathbf{x}_t$ for an individual subject on the basis of its specific data sequence $\mathbf{y}_\circ = [\mathbf{y}_{t_0}'\ \mathbf{y}_{t_0+1}'\ \cdots\ \mathbf{y}_{t_0+T-1}']'$ and the SEM parameter estimates $\widehat{\boldsymbol{\theta}}_r$ in the r^{th} iteration of the EM algorithm. The state estimates are utilized to fill in the missing data points in the data set. The completed data in turn are used to calculate an adjusted augmented moment matrix to be entered in the next M-step.

For Kalman smoothing on the basis of the structured means instead of the zero means SEM model, adjustments have to be made for the Kalman filter part of the smoother algorithm, adding fixed inputs $\mathbf{B}_{t-1}\mathbf{u}_{t-1}$ and $\mathbf{D}_t\mathbf{u}_t$ to the filter Equations 3.10 and 3.12 only. The filter consists of a measurement update, $\hat{\mathbf{x}}_t$ with associated error covariance matrix $\mathbf{P}_t$

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t-} + \mathbf{H}_t(\mathbf{y}_t - \hat{\mathbf{y}}_{t-}) \quad \text{with} \quad \hat{\mathbf{y}}_{t-} = \mathbf{C}_t\hat{\mathbf{x}}_{t-} + \mathbf{D}_t\mathbf{u}_t\ , \tag{3.10}$$

$$\mathbf{P}_t = \mathbf{M}_t\mathbf{P}_{t-}\mathbf{M}_t' + \mathbf{H}_t\mathbf{R}_t\mathbf{H}_t'\ , \tag{3.11}$$

for filter error $\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$, and a time update, $\hat{\mathbf{x}}_{t-}$ with associated error covariance matrix $\mathbf{P}_{t-}$

$$\hat{\mathbf{x}}_{t-} = \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1}\ , \tag{3.12}$$

$$\mathbf{P}_{t-} = \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}' + \mathbf{Q}_{t-1}\ . \tag{3.13}$$

$\mathbf{M}_t = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)$ while $\mathbf{H}_t = \mathbf{P}_{t-}\mathbf{C}_t'(\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t)^{-1}$ is the Kalman gain matrix. The state and output predictors $\hat{\mathbf{x}}_{t-}$ and $\hat{\mathbf{y}}_{t-}$ utilize past information, whereas the measurement update $\hat{\mathbf{x}}_t$ additionally processes current observations in $\mathbf{y}_t$.

At the initial time point $t = t_0$ where the first data are present, no past information is available. By substituting the unconditional mean $E(\mathbf{x}_{t_0})$ for $\hat{\mathbf{x}}_{t_0-}$ and the unconditional covariance matrix $cov(\mathbf{x}_{t_0})$ for $\mathbf{P}_{t_0-}$, both obtainable from the SEM output, the Kalman filter in Equations 3.10 and 3.11 leads to an estimator which is optimal for information at the single time point t_0 (Singer, 1992, p. 141). It is also known as the regression estimator in factor analysis (Lawley & Maxwell, 1971).

The adjustments for the input-variables in the Kalman filter result in the correct smoother state estimates, to be used in the E-step of the EM algorithm. It means that the backward smoother formulation is the same for the structured means as for the zero means process (Jansen & Oud, 1995; Rauch et al., 1965).

As is shown by Jansen and Oud (1995), subject specific moment matrices need to be computed on the basis of the smoother state estimates, using the parameter estimates $\widehat{\boldsymbol{\theta}}_r$ from a previous iteration. As $\mathbf{S}$ can be written as a sum of subject specific matrices $\mathbf{S}_i$:

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{S}_i = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\, \mathbf{y}_i' ,$$

only the $\mathbf{S}_i$ of subjects with missing data, and only at the places where missing data occur, need to be changed to $\mathbf{S}_{i,r+1}$ for the calculation of $\mathbf{S}_{r+1}$. Here also, few adjustments are needed for the computation of subject specific moment matrices. In each $\mathbf{S}_{i,r+1}$ the missing blocks $\mathbf{y}_{it}\, \mathbf{y}_{it}'$ and $\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}'$ are filled out with conditional expectations

$$E(\mathbf{y}_{it}\, \mathbf{y}_{it}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) = \hat{\mathbf{y}}_{itr}^{s}\, \hat{\mathbf{y}}_{itr}^{s\prime} + \mathbf{C}_{tr}\mathbf{P}_{itr}^{s}\mathbf{C}_{tr}' + \mathbf{R}_{tr} , \tag{3.14}$$

and

$$E(\mathbf{y}_{it}\, \mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) = \hat{\mathbf{y}}_{itr}^{s}\, \hat{\mathbf{y}}_{i,t-k,r}^{s\prime} + \mathbf{C}_{tr}\mathbf{P}_{it,t-k,r}^{s}\mathbf{C}_{t-k,r}' . \tag{3.15}$$

The $\hat{\mathbf{y}}_{itr}^{s} = \mathbf{C}_{tr}\hat{\mathbf{x}}_{itr}^{s} + \mathbf{D}_{tr}\mathbf{u}_t$ represent the estimates (conditional expectations) of the subject specific missing data points. $\mathbf{D}_{tr}\mathbf{u}_t$ is seen to accommodate the modifications for the structured means process. $\hat{\mathbf{x}}_{itr}^{s}$ is the vector of the smoother state estimates, while the $\mathbf{P}_{itr}^{s}$ represent the diagonal blocks of the smoother estimation error covariance matrix. The role of $\mathbf{P}_{it,t-k,r}^{s}$ for the off-diagonal blocks is analogous to $\mathbf{P}_{itr}^{s}$ for the diagonal blocks (see Jansen & Oud, 1995). The blocks combining observed $\mathbf{y}_{it,obs}$ with missing $\mathbf{y}_{i,t-k}$, missing $\mathbf{y}_{it}$ with observed $\mathbf{y}_{i,t-k,obs}$, and fixed observed $\mathbf{u}_i$ with missing $\mathbf{y}_{it}$ become, respectively:

$$\begin{aligned}
E(\mathbf{y}_{it,obs}\, \mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) &= \mathbf{y}_{it,obs}\, E(\mathbf{y}_{i,t-k}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) \\
&= \mathbf{y}_{it,obs}\, \hat{\mathbf{y}}_{i,t-k\,r}^{s\prime} , \qquad (3.16) \\
E(\mathbf{y}_{it}\, \mathbf{y}_{i,t-k,obs}' | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r) &= E(\mathbf{y}_{it} | \mathbf{y}_{i,obs}, \widehat{\boldsymbol{\theta}}_r)\, \mathbf{y}_{i,t-k,obs}'
\end{aligned}$$

$$
\begin{aligned}
&= \hat{\mathbf{y}}^{s}_{itr}\ \mathbf{y}'_{i,t-k,obs}\ , & (3.17)\\
E(\mathbf{u}_i\ \mathbf{y}'_{it}|\mathbf{y}_{i,obs},\widehat{\boldsymbol{\theta}}_r) &= \mathbf{u}_i\ E(\mathbf{y}'_{it}|\mathbf{y}_{i,obs},\widehat{\boldsymbol{\theta}}_r) \\
&= \mathbf{u}_i\ \hat{\mathbf{y}}^{s\prime}_{itr}\ . & (3.18)
\end{aligned}
$$

In the next section, it is shown how the procedure outlined above is used in the construction of the pupil monitoring system LISKAL.

3.5 Educational research example of decoding speed

The missing data procedure was applied to a longitudinal SEM model describing the latent mean development of decoding speed for a sample of $N = 1412$ pupils, drawn from the Dutch population of primary school pupils. Tests on the ability of decoding as many words as possible from a list within a one minute period were taken two times a year over a period of three years with a total of six observation time points. At each observation time point two test forms, A and B (Brus & Voeten, 1979) were used. The data set covered a period from grade 4 to grade 6 at the primary school level. The distribution of missing data over the total of twelve variables is shown in Table 3.1. For $n = 674$ pupils no missing observations appeared. The total percentage of missings over $N \times p \times T$ observations was 32.3%. It was assumed that the data were missing at random or MAR (Little & Rubin, 1987; Rubin, 1976).

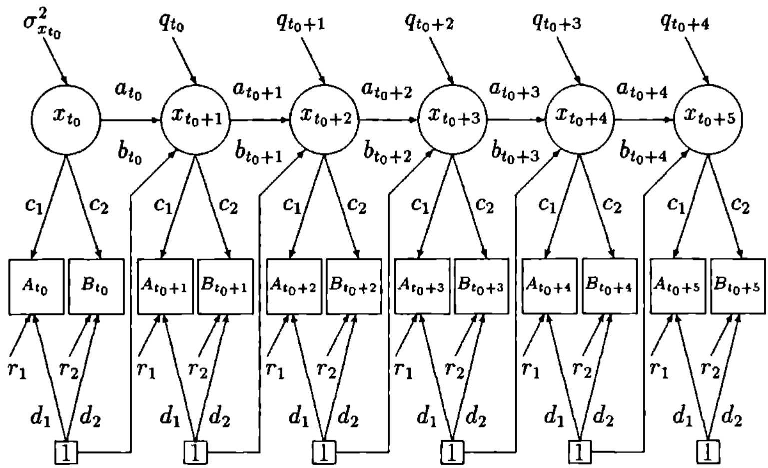
The SEM path diagram is shown in Figure 3.1. It is a first-order autoregressive scheme with the circles representing the latent decoding speed variables x_t. Because the model accounts for the measurement errors r_1 and r_2 in the observables A_t and B_t, the estimates of the autoregressive coefficients a_t and latent intercept terms b_t are disattenuated for unreliability. The small squares contain the unit input-variable which can be seen to affect both the latent states and output variables. Because it is constant over time and common to all sample units, it only needs to be specified once in the SEM model.

Because the same pair of tests was used over time, the measurement parameters were constrained to be time-invariant; constant loadings c_1 and c_2, measurement origins d_1 and d_2, and measurement error variances r_1 and r_2. Identification of the

Tab. 3.1: Number and distribution of the missing observations, and total percentage of missing observations in the sample of $N = 1412$ units.

No. of missing observations	0	1	2	3	4	5	6	7	8	9	10
No. of cases	674	7	68	0	83	1	28	0	341	4	206
Percentage of missing observations = 32.3%											

Fig. 3.1: Structured means SEM model for decoding speed.



model was obtained by fixing d_1 to zero and c_1 to unity. As a result, the initial latent mean $\mu_{x_{t_0}}$ equaled the observed mean $\mu_{A_{t_0}}$, and the initial latent variance equaled the initial observed variance minus its measurement error variance; $\sigma^2_{x_{t_0}} = \sigma^2_{A_{t_0}} - r_1$. In this way the latent decoding speed development was defined in terms of the scale units and measurement origins of the first instrument A_{t_0}. Except for the initial latent variance $\sigma^2_{x_{t_0}}$, the latent and observed means and variances are not displayed in Figure 3.1. Altogether 21 parameters had to be estimated on the basis of a moment matrix of 13 observed variables (including the unit input-variable), leaving 70 degrees of freedom in the model.

3.6 Results and discussion

The results of the EM missing data procedure are shown in Table 3.2. A total of 16 EM iterations were needed for the parameter solution to converge, meaning that additional iterations did not change the SEM parameter estimates. The initial estimates were based on a SEM analysis of the subsample of $n = 674$ pupils having complete data. Table 3.2 shows that all the autoregressive coefficients are above .91, indicating that pupils tend to keep almost the same achievement level over time; pupils with high or low scores at one time point have high or low scores at other time points as well. The latent mean values are based on the latent intercept terms b_t and autoregressive coefficients a_t, and can be calculated by means of Equation 3.19 (Jöreskog & Sörbom, 1985).

Tab. 3.2: Converged parameter estimates and standard errors for the structured means SEM model of decoding speed, after application of EM algorithm.

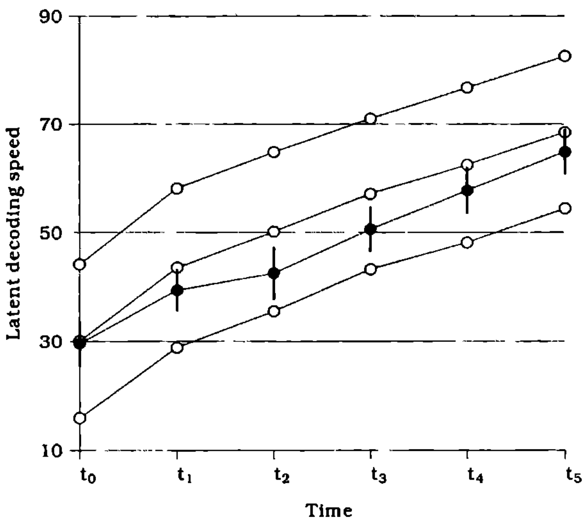
	Coefficients	Estimates	s.e.
Autoregressive coefficients	a_{t_0}	.978	(.012)
	a_{t_0+1}	.953	(.010)
	a_{t_0+2}	.915	(.009)
	a_{t_0+3}	.998	(.009)
	a_{t_0+4}	.965	(.008)
Initial latent variance	$\sigma^2_{x_{t_0}}$	195.25	(7.57)
Process error variances	q_{t_0}	24.90	(1.31)
	q_{t_0+1}	20.91	(1.12)
	q_{t_0+2}	11.84	(.75)
	q_{t_0+3}	12.33	(.78)
	q_{t_0+4}	8.45	(.68)
Initial latent mean	$\mu_{x_{t_0}}$	30.07	(.38)
and intercepts	b_{t_0}	13.86	(.38)
	b_{t_0+1}	8.77	(.47)
	b_{t_0+2}	11.46	(.45)
	b_{t_0+3}	5.39	(.55)
	b_{t_0+4}	8.26	(.53)
Latent means	$\mu_{x_{t_0}}$	30.07	
	$\mu_{x_{t_0+1}}$	43.28	
	$\mu_{x_{t_0+2}}$	50.04	
	$\mu_{x_{t_0+3}}$	57.27	
	$\mu_{x_{t_0+4}}$	62.54	
	$\mu_{x_{t_0+5}}$	68.62	
Factorloadings	c_1	1.000	
	c_2	1.014	(.003)
Measurement origins	d_1	0.000	
	d_2	-1.792	(.151)
Measurement error variances	r_1	10.93	(0.27)
	r_2	11.00	(0.27)
Number of EM iterations = 16			

$$E(x_t) = a_t E(x_{t-1}) + b_{t-1} \ . \tag{3.19}$$

The latent means as well as the latent standard deviations are easily obtainable from the SEM output. Figure 3.2 displays the estimated latent mean developmental curve, plus and minus one standard deviation (lines with white circles), and an individual developmental curve (line with black circles), plus and minus the standard errors of estimation (vertical lines). The population mean developmental curve shows an absolute increase of the ability of decoding speed over time. The individual growth curve starts just below the mean curve at the initial point in time and remains below it. Relatively, the pupil's achievement decreases over the first two periods. On an absolute scale, however, the pupil's ability increases as time proceeds. From the third time point onwards, the pupils' ability level approaches the mean curve, implying that the individual absolute growth is somewhat higher than the mean absolute growth.

The missing data procedure is very general. It can be applied to all recursive longitudinal SEM models, including higher order ARMA models (Oud et al., 1993). An advantage compared to other types of methods (for an overview see Little & Rubin, 1987, chap. 4.) is that all available data is processed and that the correct maximum likelihood estimates are obtained. Moreover, the moment matrices, computed in each E-step, are kept Gramian, which prevents various kinds of improper solutions (Bentler & Jamshidian, 1994)

Fig. 3.2: Latent mean development ± the standard deviations (white circles) for decoding speed based on a structured means SEM model; individual growth curve (black circles) ± the standard errors of estimation (vertical lines).



Chapter 4

Nonstandard constraints in SEM state space analysis of panel data: modeling stationarity and overlapping cohort commonness [1]

Abstract

Nonstandard linear and nonlinear constraints are applied in SEM state space analysis for modeling first-order and second-order stationary processes, and for modeling commonness (common model implied means and/or variances-covariances) in cohorts of the overlapping cohort design by means of a multi-sample SEM analysis. General matrix algebraic constraint formulas are derived which restrict the latent means and covariance functions over time and between different cohorts. The implications of stationarity with regard to the time-varying properties of the model parameters are clarified. The specification and testing of commonness in successive cohorts is pointed out to solve the initialization problem in the overlapping cohort design. Finally, the implications of stationarity for commonness in an overlapping cohort analysis are treated.

4.1 Introduction

SEM modeling is frequently employed in the analysis of panel data. In addition to models with polynomial random effects (Meredith & Tisak, 1990; Rogosa & Willet, 1985; Willett & Sayer, 1994), the causally dynamically orientated SSM has been used in SEM panel data analysis (MacCallum & Ashby, 1986; Oud,

[1] This article has been submitted Jansen, R A R G , & Oud, J H L (1996) Nonstandard constraints in SEM state space analysis of panel data modeling stationarity and overlapping cohort commonness

van den Bercken & Essers, 1990). Originating from system and control theory, the SSM became popular in time series analysis, econometrics and related areas (Caines, 1988). Application of the SSM allows the use of optimal filtering and smoothing techniques for estimating latent characteristics of individual subjects as well as optimal control procedures for controlling output behavior (Jazwinski, 1970; Rauch, Tung & Striebel, 1965). The generality of the SSM stems from the fact that for each autoregressive moving-average (ARMA) structure of arbitrary order, whether time-invariant or time-varying, a SSM can be found reproducing its behavior (Caines, 1988, p. 111). For implementing higher-order ARMA structures by means of SEM see Oud and Jansen (1995).

SEM software packages traditionally offer two kinds of standard constraints: fixing parameters at specific values and restricting different parameters to be equal (e.g. Jöreskog & Sörbom, 1989, p. 13). The use of indirect techniques, for example, employing so-called phantom or imaginary variables (Rindskopf, 1984), gave some opportunity to extend the class of standard constraints. Nowadays, several packages offer the possibility of applying nonstandard linear and nonlinear constraints directly (see Dolan & Molenaar, 1993). While often constraints can only be specified in scalar form in terms of single parameters (e.g. Cardon, Fulker, & Jöreskog, 1991), Neale's (1995) Mx software package, used in this paper, allows standard and nonstandard constraints to be formulated using matrix algebraic expressions.

Processes modeled by the SSM may be second-order stationary (constant mean trajectories and covariance functions depending on the time interval only), first-order stationary (constant mean trajectories only), or nonstationary. Nonstandard linear and nonlinear constraints are applied in two related fields of SEM-SSM modeling. First, for the implementation of stationarity. Second, in a multi-sample SEM analysis of the overlapping cohort design (OCD). Here constraints are applied for testing and estimating commonness (common model implied means and/or variances-covariances) of the partly overlapping cohorts. While (second-order) stationarity is based on parameters specified invariant over time, commonness is to be specified by parameters invariant over the cohorts.

4.2 Longitudinal SEM model in state space form

The linear stochastic discrete time SSM consists of two equations. The dynamic part or state transition equation describing the latent state's development over time, and the static part or measurement equation describing what is actually observed.

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{G}_{t-1}\mathbf{w}_{t-1} \; . \tag{4.1}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{H}_t\mathbf{v}_t \; . \tag{4.2}$$

$\mathbf{x}_t$ represents the m-vector of the state variables at time t. $m \times m$ matrix $\mathbf{A}_{t-1}$ contains the autoregressive and cross-lagged effects describing how the state at time point $t-1$ changes into the state at the next time point t ($t, t-1 \in \{t_0, t_0+1, \ldots, t_0+T-1\}$ where integers t_0 and $T \geq 2$ are, respectively, the initial time point and the total number of time points considered). Vectors $\mathbf{u}_{t-1}$ and $\mathbf{u}_t$ contain the fixed input-variables, and $\mathbf{B}_{t-1}$ and $\mathbf{D}_t$ their effects. $\mathbf{w}_{t-1}$ is a vector of standardized errors with mean vector $E(\mathbf{w}_{t-1}) = \mathbf{0}$ and covariance matrix $cov(\mathbf{w}_{t-1}) = \mathbf{I}$. The advantage of the parameterization of the process error covariance matrix $\mathbf{Q}_{t-1} \equiv cov(\mathbf{G}_{t-1}\mathbf{w}_{t-1}) = \mathbf{G}_{t-1}\mathbf{G}'_{t-1}$ in terms of $\mathbf{G}_{t-1}$ is that negative variance estimates are avoided by the SEM program. The state equation may be modeled in SEM such that $\mathbf{G}_{t-1}$ and thus $\mathbf{Q}_{t-1}$ are diagonal. $\mathbf{G}_{t-1}$ may also be modeled as part of the reduced form of a recursive structural equation ($\mathbf{G}_{t-1}$ triangular) or an interdependent structural equation ($\mathbf{G}_{t-1}$ nontriangular).

The $p \times m$ matrix $\mathbf{C}_t$ defines how the unobserved state variables are related to the observables in $\mathbf{y}_t$. $\mathbf{v}_t$ is another vector of standardized errors, having $E(\mathbf{v}_t) = \mathbf{0}$ and $cov(\mathbf{v}_t) = \mathbf{I}$. The measurement error covariance matrix $\mathbf{R}_t \equiv cov(\mathbf{H}_t\mathbf{v}_t) = \mathbf{H}_t\mathbf{H}'_t$ relates in the same way to $\mathbf{H}_t$ as the process error covariance matrix $\mathbf{Q}_{t-1}$ does to $\mathbf{G}_{t-1}$.

It is assumed that the errors in different vectors $\mathbf{w}_t$ and $\mathbf{v}_t$ have zero covariances: $E(\mathbf{w}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$, $E(\mathbf{w}_t\ \mathbf{w}'_{t'}) = \mathbf{I}_t\delta_{t-t'}$, $E(\mathbf{v}_t\ \mathbf{v}'_{t'}) = \mathbf{I}_t\delta_{t-t'}$ for all t and t' ($\delta_{t-t'}$ Kronecker's delta, being 0 if $t \neq t'$ and 1 if $t = t'$), and have zero covariances with the initial state: $E(\mathbf{w}_t\ \mathbf{x}'_{t_0}) = E(\mathbf{v}_t\ \mathbf{x}'_{t_0}) = \mathbf{0}$ for all t. Finally it is assumed that the error vectors and the initial state are jointly multinormally distributed.

A process $\mathbf{x}_t$ is said to be stationary up to order n if, for any admissible set of time points $t_1, t_2, ..., t_j$ and time interval k, all the joint moments up to order n of $\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, ..., \mathbf{x}_{t_j}$ exist and equal the corresponding joint moments up to order n of $\mathbf{x}_{t_1+k}, \mathbf{x}_{t_2+k}, ..., \mathbf{x}_{t_j+k}$ (Priestly, 1981, p. 105-106). Therefore, first-order stationarity is defined by $E(\mathbf{x}_t) = E(\mathbf{x}_{t+k})$ for all t and k (constant mean trajectories), and second-order stationarity by the additional condition $\mathbf{\Phi}_{t,t'} = \mathbf{\Phi}_{t+k,t'+k} = \mathbf{\Phi}_{t-t'}$ for all t, t' and k, where $\mathbf{\Phi}_{t,t'} = E([\mathbf{x}_t - E(\mathbf{x}_t)][\mathbf{x}_{t'} - E(\mathbf{x}_{t'})]')$ (covariance function depending on the time interval $t - t'$ only).

According to the above definition second-order stationarity implies first-order stationarity, where the first-order stationary process may be a zero-means or nonzero-means process. This gives four types of stationary processes to be treated in this paper: a first-order zero-means and a first-order nonzero-means stationary process, a second-order stationary process for either a first-order zero-means or a first-order nonzero-means stationary process. Stationarity can be applied to the observed and latent variables processes. Our focus is on the the latent variables processes which are easily transformed into observed variables processes of the same type.

By specifying $\mathbf{B}_{t-1}\mathbf{u}_{t-1} = \mathbf{D}_t\mathbf{u}_t = \mathbf{0}$ and $E(\mathbf{x}_{t_0}) = E(\mathbf{y}_{t_0}) = \mathbf{0}$, implying

$E(\mathbf{x}_t) = E(\mathbf{y}_t) = \mathbf{0}$, the SSM becomes first-order zero-means stationary. It allows the model to be kept second-order nonstationary, however, meaning that $E(\mathbf{x}_t\,\mathbf{x}'_{t'})$ and $E(\mathbf{y}_t\,\mathbf{y}'_{t'})$ may vary across time even for equal intervals $t - t'$. In dropping the assumption of first-order zero-means stationarity, specifying $\mathbf{B}_{t-1}\mathbf{u}_{t-1} \neq \mathbf{0}$, $\mathbf{D}_t\mathbf{u}_t \neq \mathbf{0}$, and possibly $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, the nonzero-means of the latent and observed variables are modeled as follows:

$$E(\mathbf{x}_t) = \mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k\ , \tag{4.3}$$

$$E(\mathbf{y}_t) = \mathbf{C}_t\mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \mathbf{C}_t\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k + \mathbf{D}_t\mathbf{u}_t\ . \tag{4.4}$$

$\mathcal{A}_{t,t_0} = \prod_{k=1}^{t-t_0}\mathbf{A}_{t-k}$ is the well-known state transition matrix, also defined for $t = t_0$: $\mathcal{A}_{t_0,t_0} = \mathcal{A}_{t,t} \equiv \mathbf{I}$ (Desoer, 1970, p. 71).

In one case only a single so-called unit input-variable is specified ($u_t = 1$ for all t), which is constant over time as well as over subjects in the sample. Here vectors $\mathbf{b}_{t-1}$ represent the latent growth intercepts and vectors $\mathbf{d}_t$ the location parameters (origins) of the measurement instruments. The model implies a means process which is common to all subjects in the sample. In another case the input-variables are all constant over time ($\mathbf{u}_t = \mathbf{u}_{t-k}$ for all t and $k > 0$) but, apart from the unit input-variable, varying over subjects (e.g. background variables like gender or socioeconomic status). Finally, additional input-variables may be specified that vary over time points as well as over subjects.

Whether $E(\mathbf{x}_{t_0}) = \mathbf{0}$ or $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, the initial state mean $E(\mathbf{x}_{t_0})$ may be modeled by means of an additional matrix $\mathbf{B}_{t_0-1}$, to be specified zero except, in the case of $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, for the elements corresponding to the unit input-variable in $\mathbf{u}_{t_0-1}$:

$$E(\mathbf{x}_{t_0}) = \mathbf{B}_{t_0-1}\mathbf{u}_{t_0-1}\ , \tag{4.5}$$

$$E(\mathbf{y}_{t_0}) = \mathbf{C}_{t_0}E(\mathbf{x}_{t_0}) + \mathbf{D}_{t_0}\mathbf{u}_{t_0}\ , \tag{4.6}$$

In deriving the SEM model, first write Equations 4.1 and 4.2 as follows:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{t-1} & \mathbf{B}_{t-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{u}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{t-1}\mathbf{w}_{t-1} \\ \mathbf{u}_t \end{bmatrix}, \tag{4.7}$$

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{C}_t & \mathbf{D}_t \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} + \begin{bmatrix} \mathbf{H}_t\mathbf{v}_t \\ \mathbf{0} \end{bmatrix}. \tag{4.8}$$

Then, collecting all input-variables in the input-vector $\mathbf{u}$ but specifying the constant (e.g. the unit input-variable) and other exactly linearly related input-variables only once in $\mathbf{u}$, and defining

$$
\begin{array}{lll}
\boldsymbol{\eta} = [\mathbf{x}'\ \mathbf{u}']' & \text{with} & \mathbf{x} = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1} \cdots \mathbf{x}'_{t_0+T-1}]', \\
\mathbf{y} = [\mathbf{y}'_o\ \mathbf{u}']' & \text{with} & \mathbf{y}_o = [\mathbf{y}'_{t_0}\ \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_{t_0+T-1}]', \\
\boldsymbol{\zeta} = [(\overline{\mathbf{G}}\mathbf{w})'\ \mathbf{u}']' & \text{with} & \overline{\mathbf{G}}\mathbf{w} = [(\mathbf{G}_{t_0-1}\mathbf{w}_{t_0-1})'\ (\mathbf{G}_{t_0}\mathbf{w}_{t_0})' \cdots (\mathbf{G}_{t_0+T-2}\mathbf{w}_{t_0+T-2})']', \\
\boldsymbol{\varepsilon} = [(\overline{\mathbf{H}}\mathbf{v})'\ \mathbf{0}']' & \text{with} & \overline{\mathbf{H}}\mathbf{v} = [(\mathbf{H}_{t_0}\mathbf{v}_{t_0})'(\mathbf{H}_{t_0+1}\mathbf{v}_{t_0+1})' \cdots (\mathbf{H}_{t_0+T-1}\mathbf{v}_{t_0+T-1})']',
\end{array}
$$

the SEM model is derived:

$$
\begin{array}{ccccccc}
\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} & = & \begin{bmatrix} \overline{\mathbf{A}} & \overline{\mathbf{B}} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} & + & \begin{bmatrix} \overline{\mathbf{G}}\mathbf{w} \\ \mathbf{u} \end{bmatrix} & , \\
\boldsymbol{\eta} & = & \boldsymbol{B} & \boldsymbol{\eta} & + & \boldsymbol{\zeta} & \text{with} \quad cov(\boldsymbol{\zeta}) = \boldsymbol{\Psi}
\end{array}
\tag{4.9}
$$

$$
\begin{array}{ccccccc}
\begin{bmatrix} \mathbf{y}_o \\ \mathbf{u} \end{bmatrix} & = & \begin{bmatrix} \overline{\mathbf{C}} & \overline{\mathbf{D}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} & \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} & + & \begin{bmatrix} \overline{\mathbf{H}}\mathbf{v} \\ \mathbf{0} \end{bmatrix} & \\
\mathbf{y} & = & \boldsymbol{\Lambda} & \boldsymbol{\eta} & + & \boldsymbol{\varepsilon} & \text{with} \quad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}
\end{array}
\tag{4.10}
$$

where all parameter matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, $\mathbf{C}_t$, $\mathbf{D}_t$ are put on the appropriate places in $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, $\overline{\mathbf{C}}$, $\overline{\mathbf{D}}$, respectively. Notice that in $\mathbf{x}$ the initial state $\mathbf{x}_{t_0}$ gets zero rows in $\overline{\mathbf{A}}$ but $\mathbf{B}_{t_0-1}$ in $\overline{\mathbf{B}}$ for modeling its mean $E(\mathbf{x}_{t_0})$, which therefore is subtracted from $\mathbf{x}_{t_0}$ so that $\mathbf{G}_{t_0-1}\mathbf{w}_{t_0-1} = \mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})$. The upper left corner of $\boldsymbol{\Psi}$ contains the initial latent covariance matrix $\boldsymbol{\Phi}_{t_0} = cov(\mathbf{x}_{t_0})$.

For estimation of the SEM model the maximum likelihood fit function

$$
F_{ML} = \log \mid \boldsymbol{\Sigma} \mid + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log \mid \mathbf{S} \mid - p \ , \tag{4.11}
$$

is applied in the SEM program. Where

$$
\boldsymbol{\Sigma} \equiv E(\mathbf{y}\ \mathbf{y}') = \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \ ,
$$

on the basis of the sample moment matrix

$$
\mathbf{S}_{(p\times p)} = \tfrac{1}{N}\mathbf{Y}\ \mathbf{Y}' = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_{0i}\ \mathbf{y}'_{0i} & \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_{0i}\ \mathbf{u}'_i \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{u}_i\ \mathbf{y}'_{0i} & \frac{1}{N}\sum_{i=1}^{N} \mathbf{u}_i\ \mathbf{u}'_i \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{y_0} & \mathbf{S}_{y_0 u} \\ \mathbf{S}_{u y_0} & \mathbf{S}_u \end{bmatrix} ,
$$

with $p = p_0 + q$ and q the number of (fixed) elements in $\mathbf{u}$ (the set of $N' \leq N$ distinct fixed values $\mathbf{u}_i$ in the sample must contain at least q linearly independent ones).

It should be noted that $\overline{\mathbf{G}}\ \overline{\mathbf{G}}'$ in $\boldsymbol{\Psi}$ and $\overline{\mathbf{H}}\ \overline{\mathbf{H}}'$ in $\boldsymbol{\Theta}$ (Equations 4.9 and 4.10) imply nonlinear constraints between parameters in $\boldsymbol{\Psi}$ and $\overline{\mathbf{G}}$, and $\boldsymbol{\Theta}$ and $\overline{\mathbf{H}}$, respectively. These constraints which cause negative estimates of the variances in Ψ and Θ to be avoided, require a rather complicated reparametrization when only traditional standard equality constraints are available (Jöreskog & Sörbom, 1989, p. 239-240, 344), but can easily and directly be implemented by means of the nonstandard constraints offered by a program like Mx.

4.3 First-order stationarity

The first-order zero-means stationary SSM leads to the standard covariance structure SEM as presented, for example, by Jöreskog and Sörbom (1989, p. 1) and does not require any nonstandard constraints. In both the zero-means and nonzero-means first-order stationary model the constant observed means processes are assumed to be generated by constant latent means processes. Thus, the means of the latent state vector $\mathbf{x}_t$ at time point t, expressed as

$$E(\mathbf{x}_t) = \mathbf{A}_{t-1}E(\mathbf{x}_{t-1}) + \mathbf{B}_{t-1}\mathbf{u}_{t-1} \ , \tag{4.12}$$

or more succinctly for all time points simultaneously as $E(\mathbf{x}) = \overline{\mathbf{A}}\ E(\mathbf{x}) + \overline{\mathbf{B}}\mathbf{u}$ (see Equation 4.9) or

$$E(\mathbf{x}) = (\mathbf{I} - \overline{\mathbf{A}})^{-1}\overline{\mathbf{B}}\mathbf{u} \ , \tag{4.13}$$

must obey $E(\mathbf{x}_t) = E(\mathbf{x}_{t-1})$ for all t. The constraints may be expressed as $E(\mathbf{x}) = \mathbf{N}E(\mathbf{x})$ with $\mathbf{N}$ a square shift matrix consisting of zero and identity matrices of order m, shifting each subvector $E(\mathbf{x}_t)$ in $E(\mathbf{x})$ one subvector position downwards:

$$\mathbf{N} = \begin{bmatrix} \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \ddots & & & \mathbf{0} \\ \mathbf{0} & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} .$$

Then first-order stationarity is realized by constraints formula

$$[(\mathbf{I} - \overline{\mathbf{A}})^{-1} - \mathbf{N}(\mathbf{I} - \overline{\mathbf{A}})^{-1}]\overline{\mathbf{B}}\mathbf{u} = \mathbf{0} \ , \tag{4.14}$$

which has to be specified in SEM for every value of $\mathbf{u}$ being present in the sample. This can be done in a multi-sample SEM analysis by specifying Equation 4.14 separately for every group of individuals with the same input-value $\mathbf{u}$. A sufficient condition for first-order stationarity is

$$(\mathbf{I} - \mathbf{N})(\mathbf{I} - \overline{\mathbf{A}})^{-1}\overline{\mathbf{B}} = \mathbf{0} \ . \tag{4.15}$$

For the input being the unit input-variable ($u = 1$), Equation 4.14 becomes $(\mathbf{I} - \mathbf{N})(\mathbf{I} - \overline{\mathbf{A}})^{-1}\overline{\mathbf{b}} = \mathbf{0}$.

4.4 Second-order stationarity

A second-order stationary latent process implies that the covariance matrix $\mathbf{\Phi}_x = cov(\mathbf{x}) = (\mathbf{I} - \overline{\mathbf{A}})^{-1}\overline{\mathbf{G}}\ \overline{\mathbf{G}}'(\mathbf{I} - \overline{\mathbf{A}}')^{-1}$ has a so-called *Toeplitz* structure (Caines, 1988, p. 12). $\mathbf{\Phi}_x$ must be invariant with respect to shifts in the direction of its main diagonal. For example, $\mathbf{\Phi}_{t_0} = cov(\mathbf{x}_{t_0})$ representing the initial latent covariance matrix in $\mathbf{\Phi}_x$, equals $\mathbf{\Phi}_{t_0+1} = cov(\mathbf{x}_{t_0+1})$, which equals $\mathbf{\Phi}_{t_0+2} = cov(\mathbf{x}_{t_0+2})$, etc. But also, $\mathbf{\Phi}_{t_0+1,t_0} = \mathbf{\Phi}_{t_0+2,t_0+1} = \ldots$. Suppose that $\mathbf{\Phi}_x$ consists of 2×2 submatrices:

$$\mathbf{\Phi}_x = \begin{bmatrix} \mathbf{\Phi}_{t_0} & \mathbf{\Phi}'_{t_0+1,t_0} & \mathbf{\Phi}'_{t_0+2,t_0} \\ \mathbf{\Phi}_{t_0+1,t_0} & \mathbf{\Phi}_{t_0+1} & \mathbf{\Phi}'_{t_0+2,t_0+1} \\ \mathbf{\Phi}_{t_0+2,t_0} & \mathbf{\Phi}_{t_0+2,t_0+1} & \mathbf{\Phi}_{t_0+2} \end{bmatrix} =$$

$$\begin{bmatrix} \varphi_{11} & \varphi'_{21} & \varphi'_{31} & \varphi'_{41} & \varphi'_{51} & \varphi'_{61} \\ \varphi_{21} & \varphi_{22} & \varphi'_{32} & \varphi'_{42} & \varphi'_{52} & \varphi'_{62} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi'_{43} & \varphi'_{53} & \varphi'_{63} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} & \varphi'_{54} & \varphi'_{64} \\ \varphi_{51} & \varphi_{52} & \varphi_{53} & \varphi_{54} & \varphi_{55} & \varphi'_{65} \\ \varphi_{61} & \varphi_{62} & \varphi_{63} & \varphi_{64} & \varphi_{65} & \varphi_{66} \end{bmatrix} .$$

Then second-order stationarity is implemented by subtracting the upper-left and lower-right 4×4 submatrices of the 6×6 matrix $\mathbf{\Phi}_x$ and constraining the result to be zero. The constraints formula is as follows,

$$\mathbf{S}_1\mathbf{\Phi}_x\mathbf{S}'_1 - \mathbf{S}_2\mathbf{\Phi}_x\mathbf{S}'_2 = \mathbf{0} . \tag{4.16}$$

$\mathbf{S}_1$ and $\mathbf{S}_2$ both are rectangular selection matrices, consisting of zero and identity matrices of order m, and have the following form,

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \ldots & \ldots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \ldots & \mathbf{0} & \mathbf{I} \end{bmatrix} .$$

Second-order stationarity implies time-invariance of the parameter matrices $\mathbf{A}_t$ and $\mathbf{Q}_t$, which can be proven as follows. Assume a second-order stationary process, then:

$$E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0}) = E(\mathbf{x}_{t_0+1}\ \mathbf{x}'_{t_0+1}) = E(\mathbf{x}_{t_0+2}\ \mathbf{x}'_{t_0+2}) = \ldots \tag{4.17}$$

(4.18)

$$\begin{aligned} E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0+1}) &= E(\mathbf{x}_{t_0+1}\ \mathbf{x}'_{t_0+2}) = \ldots \\ E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0+2}) &= E(\mathbf{x}_{t_0+1}\ \mathbf{x}'_{t_0+3}) = \ldots \end{aligned} \tag{4.19}$$

Because of first-order stationarity, it follows from Equation 4.17, that

$$\begin{aligned} \mathbf{\Phi}_{t_0} = \mathbf{\Phi}_{t_0+1} &= \mathbf{A}_{t_0}\mathbf{\Phi}_{t_0}\mathbf{A}'_{t_0} + \mathbf{Q}_{t_0} = \\ \mathbf{\Phi}_{t_0+2} &= \mathbf{A}_{t_0+1}\mathbf{\Phi}_{t_0+1}\mathbf{A}'_{t_0+1} + \mathbf{Q}_{t_0+1} = \ldots \end{aligned} \tag{4.20}$$

and further from Equation 4.19, that

$$\mathbf{\Phi}_{t_0}\mathbf{A}'_{t_0} = \mathbf{\Phi}_{t_0+1}\mathbf{A}'_{t_0+1} = \ldots \tag{4.21}$$

Therefore, $\mathbf{A}_{t_0} = \mathbf{A}_{t_0+1} = \ldots$ (Equation 4.21) because $\mathbf{\Phi}_{t_0} = \mathbf{\Phi}_{t_0+1} = \mathbf{\Phi}_{t_0+2} = \ldots$, and $\mathbf{Q}_{t_0} = \mathbf{Q}_{t_0+1} = \ldots$, because of Equation 4.20.

Although this time-invariance property is a necessary condition for obtaining second-order stationarity, it is not sufficient. If $\mathbf{A}_t = \mathbf{A}$ and $\mathbf{Q}_t = \mathbf{Q}$ for all t, then

$$\begin{aligned} &\mathbf{\Phi}_{t_0+1} && = \mathbf{\Phi}_{t_0+2} && = \ldots = \\ &\text{or} \\ &\mathbf{A}\mathbf{\Phi}_{t_0}\mathbf{A}' + \mathbf{Q} && = \mathbf{A}\mathbf{\Phi}_{t_0+1}\mathbf{A}' + \mathbf{Q} = \ldots \end{aligned}$$

iff additionally $\mathbf{\Phi}_{t_0} = \mathbf{\Phi}_{t_0+1} = \mathbf{\Phi}_{t_0+2} = \ldots = \mathbf{\Phi}$, so that second-order stationarity is equivalent to first-order stationarity plus time-invariance and equal covariance matrices at the main diagonal of $\mathbf{\Phi}_{\mathbf{x}}$.

Because $\mathcal{A}_{t,t_0}$ (see Equation 4.3) becomes the matrix exponential $\mathbf{A}^{t-t_0}$, $\mathbf{\Phi}_t$ propagates according to Equation 4.22

$$\mathbf{\Phi}_t = \mathbf{A}^{t-t_0}\mathbf{\Phi}_{t_0}\mathbf{A}'^{t-t_0} + \sum_{k=t_0}^{t-1} \mathbf{A}^{t-k-1}\mathbf{Q}\mathbf{A}'^{t-k-1}\ . \tag{4.22}$$

It can be seen that a second-order stationary process involves highly nonlinear restrictions on the parameters in the matrices $\mathbf{A}$, $\mathbf{Q}$ and $\mathbf{\Phi}_{t_0}$:

$$\mathbf{\Phi}_{t_0} = \mathbf{A}^{t-t_0}\mathbf{\Phi}_{t_0}\mathbf{A}'^{t-t_0} + \sum_{k=t_0}^{t-1} \mathbf{A}^{t-k-1}\mathbf{Q}\mathbf{A}'^{t-k-1}\ . \tag{4.23}$$

Because of $\mathbf{Q}$ being positive definite, $\mathbf{A}$ must be stabilizing (all eigenvalues of $\mathbf{A}$ within the unit circle of the complex plane) for Equation 4.23 to be true. Equation 4.23, in fact, reduces to

$$\mathbf{\Phi}_{t_0} = \mathbf{A}\mathbf{\Phi}_{t_0}\mathbf{A}' + \mathbf{Q}\ , \tag{4.24}$$

for $t-1 = t_0$, so that the condition $\mathbf{\Phi}_{t_0} = \mathbf{\Phi}_{t_0+1} = \ldots = \mathbf{\Phi}$ can be replaced by the constraints on the initial latent covariance matrix parameters given in

Equation 4.24. In the scalar case, Equation 4.24 becomes the well-known condition $\varphi_{t_0} = \varphi = (1 - a^2)^{-1}q$ (Priestly, 1981, p. 119).

Analogously, the constraints on the initial latent mean vector $E(\mathbf{x}_{t_0})$, implied by second-order stationarity via $\mathbf{A}_t = \mathbf{A}$ and the constancy of $\mathbf{B}_t\mathbf{u}_t$ over all $t \geq t_0$ (say $\mathbf{B}_t\mathbf{u}_t = \mathbf{C}$) (cfr. Equation 4.3), are given by

$$E(\mathbf{x}_{t_0}) = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{C} . \tag{4.25}$$

To summarize, different sets of constraints can be formulated to impose second-order stationarity in SEM-SSM analysis. Equations 4.24 and 4.25 make clear that the nonlinearities reside in the constraints to be imposed on the initializing parameters.

It should be noted that constraints formula 4.14, which realizes first-order stationarity, does not necessarily imply time-invariance. However, if $\mathbf{A}_t = \mathbf{A}$ for all t, as in the case of second-order stationarity, it is easily seen that, because of $(\mathbf{I} - \mathbf{A})E(\mathbf{x}_t) = \mathbf{C}$ for all $t \geq t_0$, first-order stationarity $(E(\mathbf{x}_{t_0}) = E(\mathbf{x}_{t_0+1}) = \ldots)$ implies $\mathbf{B}_t\mathbf{u}_t = \mathbf{B}_{t-1}\mathbf{u}_{t-1} = \mathbf{C}$ which, in case $\mathbf{u}_t$ is constant over time, is realized by a time-invariant $\mathbf{B}$.

Although above the stationarity constraints are applied to the latent processes only, the observed processes become stationary as follows. If in addition to the process $\mathbf{x}_t$ being first-order stationary, $\mathbf{C}_t$ and $\mathbf{D}_t$ are time-invariant ($\mathbf{C}_t = \mathbf{C}$ and $\mathbf{D}_t = \mathbf{D}$, and in case of $u_t = u = 1$, $\mathbf{D}$ a time-invariant vector $\mathbf{d}$), the observed process becomes first-order stationary too. Moreover, if in addition to $\mathbf{x}_t$ being second-order stationary, $\mathbf{R}_t$ is time-invariant ($\mathbf{R}_t = \mathbf{R}$), the observed process becomes second-order stationary.

4.5 The overlapping cohort design

Nonstandard linear and nonlinear constraints are also necessary in modeling the OCD adequately. Figure 4.1 shows an example of the OCD. We assume samples to be randomly drawn from different cohorts (populations defined by time of birth). Each cohort covers a particular age period whose observation time points overlap with one or more other cohorts. At the overlapping time points (3 and 4, 5 and 6, and 7 and 8) in Figure 4.1, the same variables are observed in the overlapping cohorts. The samples are observed over the same historical time period (1 $\rightarrow$ 2 $\rightarrow$ 3 $\rightarrow$ 4 in cohort 1 represents the same historical period as 3 $\rightarrow$ 4 $\rightarrow$ 5 $\rightarrow$ 6 in cohort 2, etc.).

The use of an OCD in longitudinal research can be advocated for two reasons. First, it provides a method for studying long term development and change in a relatively short period of time because it allows the data collection period to be chosen as a fraction only of the total age period covered by the model. Second, it considerably limits panel attrition; for example in Figure 4.1, the subjects of the four adjacent samples need only be questioned four times, instead of questioning the subjects of one sample ten times repeatedly. In cohort 1, for example, no attrition can take place at time points $5 - 10$.

Fig. 4.1: Overlapping cohort design.

$$
\begin{array}{lccccccccccccccccccc}
\text{Cohort 1} & 1 & \rightarrow & 2 & \rightarrow & 3 & \rightarrow & 4 & & & & & & & & & & & & \\
 & & & & & | & | & | & & & & & & & & & & & & \\
\text{Cohort 2} & & & & & 3 & \rightarrow & 4 & \rightarrow & 5 & \rightarrow & 6 & & & & & & & & \\
 & & & & & & & & & | & | & | & & & & & & & & \\
\text{Cohort 3} & & & & & & & & & 5 & \rightarrow & 6 & \rightarrow & 7 & \rightarrow & 8 & & & & \\
 & & & & & & & & & & & & & | & | & | & & & & \\
\text{Cohort 4} & & & & & & & & & & & & & 7 & \rightarrow & 8 & \rightarrow & 9 & \rightarrow & 10
\end{array}
$$

The OCD has been used in growth curve analysis (Bell, 1953, 1954; Duncan, Duncan & Hops, 1996; McArdle & Hamagami, 1991; Meredith & Tisak, 1990; Raudenbush & Chan, 1992), in hierarchical linear modeling (Raudenbush & Chan, 1993), and in SEM modeling (Horn & McArdle, 1980; Oud, van Leeuwe & Jansen, 1993). It typically involves the specification and testing of common measurement and structural characteristics for the overlapping model parts of the cohorts, as is explained below. Besides that, we treat the problem of initialization of the different samples and the implications of within-group stationarity restrictions for the commonness specification of different cohorts.

The OCD is a special case of the general model for the analysis of longitudinal data from multiple populations (Jöreskog & Sörbom, 1985). It has p variables in each cohort g measured over T historical time points. The response $y_{irt}^{(g)}$ is a function of the cohort $g = 1, 2, \ldots, G$ to which the individual belongs, an observed variable $r = 1, 2, \ldots, p$, and time point t. The model of Equations 4.1 and 4.2 is formulated for each cohort separately,

$$
\begin{aligned}
\mathbf{x}_t^{(g)} &= \mathbf{A}_{t-1}^{(g)}\mathbf{x}_{t-1}^{(g)} + \mathbf{B}_{t-1}^{(g)}\mathbf{u}_{t-1}^{(g)} + \mathbf{G}_{t-1}^{(g)}\mathbf{w}_{t-1}^{(g)} \;, && (4.26)\\
\mathbf{y}_t^{(g)} &= \mathbf{C}_t^{(g)}\mathbf{x}_t^{(g)} + \mathbf{D}_t^{(g)}\mathbf{u}_t^{(g)} + \mathbf{H}_t^{(g)}\mathbf{v}_t^{(g)} \;. && (4.27)
\end{aligned}
$$

The historical time of measurement minus the time of birth known for each cohort g, gives the age. t in Equations 4.26 and 4.27 is typically interpreted as representing age. Figure 4.1 can then be interpreted in terms of the model parameters. The arrows within each sample represent the dynamic parameters $\mathbf{A}_{t-1}^{(g)}$, $\mathbf{B}_{t-1}^{(g)}$ and $\mathbf{G}_{t-1}^{(g)}$ of the structural equation going from $t-1$ to t, whereas the numbers represent the static parameters $\mathbf{C}_t^{(g)}$, $\mathbf{D}_t^{(g)}$ and $\mathbf{H}_t^{(g)}$ of the measurement equation, and the vertical solid lines the constraints between the parameters of different cohorts.

Before it can be tested whether the OCD has common model implied latent characteristics in the form of latent mean trajectories and covariance functions, the measurement part of the OCD must obey two requirements which are assumed to be satisfied. First, the latent scale origins and units are kept identical over time, for example, by taking the same measurement instruments and equaling the $\mathbf{C}_t$ and $\mathbf{D}_t$ parameters for all t. An alternative procedure is given in Oud et al. (1993, pp. 15-16). Second, at the overlapping time points, identical latent scale origins and units are obtained by equaling the $\mathbf{C}_t^{(g)}$ and $\mathbf{D}_t^{(g)}$ parameters for different adjacent cohorts g.

4.5.1 Modeling common latent means

In modeling commonness with regard to latent nonzero means and variances-covariances, the constraints are restricted to the overlapping parts of the model. Because the constraints are applied to the results of recursive processes, however, the whole model preceding the time point of a constraint is, in fact, involved. We assume the fixed input to consist of the unit input-variable ($\mathbf{B}_t\mathbf{u}_t = \mathbf{b}_t$). The procedure is easily generalized, however, to other input-variables which vary over subjects (e.g. gender), that is, to several input defined groups, which have to be kept separate in a multi-sample SEM analysis. In this way, the total number of samples in the OCD analysis is the number of cohorts times the number of distinct input-values.

Consider a first-order nonzero-means stationary OCD. A common latent nonzero-mean trajectory over cohorts implies that $E(\mathbf{x}_{t+i})^{(g)} = E(\mathbf{x}_{t+i})^{(g-1)}$ for all $i \in \{0, \ldots, T_{g,g-1}\}$, where $T_{g,g-1}$ is the total number of overlapping time points of cohorts g and $g-1$. For $i = 0$, calling $\mathbf{x}_{t_1}^{(g)}$ the initial state vector in cohort g, and $\mathbf{x}_{t_0}^{(g-1)}$ the initial state vector in cohort $g-1$,

$$\begin{aligned} E(\mathbf{x}_{t_1})^{(g)} &= E(\mathbf{x}_{t_1})^{(g-1)} \\ &= \mathcal{A}_{t_1,t_0}^{(g-1)} E(\mathbf{x}_{t_0})^{(g-1)} + \sum_{k=t_0}^{t-1} \mathcal{A}_{t_1,k+1}^{(g-1)} \mathbf{b}_k^{(g-1)} \\ &\equiv \mathcal{U}^{(g,g-1)} , \end{aligned} \tag{4.28}$$

which for $i = 1$ results in,

$$\begin{aligned} E(\mathbf{x}_{t_1+1})^{(g)} &= E(\mathbf{x}_{t_1+1})^{(g-1)} \\ \mathbf{A}_{t_1}^{(g)} E(\mathbf{x}_{t_1})^{(g)} + \mathbf{b}_{t_1}^{(g)} &= \mathbf{A}_{t_1}^{(g-1)} E(\mathbf{x}_{t_1})^{(g-1)} + \mathbf{b}_{t_1}^{(g-1)} \\ \mathbf{A}_{t_1}^{(g)} \mathcal{U}^{(g,g-1)} + \mathbf{b}_{t_1}^{(g)} &= \mathbf{A}_{t_1}^{(g-1)} \mathcal{U}^{(g,g-1)} + \mathbf{b}_{t_1}^{(g-1)} , \end{aligned} \tag{4.29}$$

and involves the product term $\mathcal{A}_{t_1,t_0} = \prod_{k=1}^{t_1-t_0} \mathbf{A}_{t_1-k}$. Equation 4.28, in fact, solves an initialization problem in cohort g. There is no past information available for modeling $E(\mathbf{x}_{t_1})^{(g)}$, except for the information in $\mathcal{U}^{(g,g-1)}$, handed over from cohort $g-1$ to cohort g by means of a nonstandard constraint. The constraints formula for restricting the latent means at the overlapping parts of the model according to Equation 4.29 (cfr. Equation 4.13) is

$$\mathbf{N}^{(g)}[(\mathbf{I}-\overline{\mathbf{A}}^{(g)})^{-1}\overline{\mathbf{b}}^{(g)}] - \mathbf{N}^{(g-1)}[(\mathbf{I}-\overline{\mathbf{A}}^{(g-1)})^{-1}\overline{\mathbf{b}}^{(g-1)}] = \mathbf{0}\ , \tag{4.30}$$

with $\mathbf{N}^{(g)}$ and $\mathbf{N}^{(g-1)}$ both matrices consisting of zeroes except for one unit element in each row, selecting the appropriate latent mean values from $E(\mathbf{x})^{(g)}$ and $E(\mathbf{x})^{(g-1)}$, respectively. For Figure 4.1, for example, the left-hand side of Equation 4.30 could be of the form

$$\begin{array}{cccc} \mathbf{N}^{(g)} & E(\mathbf{x})^{(g)} & - \quad \mathbf{N}^{(g-1)} & E(\mathbf{x})^{(g-1)} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} & \begin{bmatrix} E(\mathbf{x}_t) \\ E(\mathbf{x}_{t+1}) \\ E(\mathbf{x}_{t+2}) \\ E(\mathbf{x}_{t+3}) \end{bmatrix}^{(g)} & - \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} & \begin{bmatrix} E(\mathbf{x}_{t-2}) \\ E(\mathbf{x}_{t-1}) \\ E(\mathbf{x}_t) \\ E(\mathbf{x}_{t+1}) \end{bmatrix}^{(g-1)} \end{array}$$

Note from Equation 4.29 that in case $\mathbf{A}_{t_1}^{(g)}$ equals $\mathbf{A}_{t_1}^{(g-1)}$, then $\mathbf{b}_{t_1}^{(g)}$ equals $\mathbf{b}_{t_1}^{(g-1)}$, but not necessarily so conversely. It should be noted that through Equations 4.28-4.29, Equation 4.30 imposes rather complex constraints on the parameter matrices $\mathbf{A}_{t-1}$ and $\mathbf{b}_{t-1}$, which cannot be imposed, specifically not as regards the initial mean $E(\mathbf{x}_{t_1})^{(g)}$, by means of standard constraints.

4.5.2 Modeling common latent variances and covariances

Commonness of latent covariance matrices means that $\mathbf{\Phi}_{t+\imath}^{(g)} = \mathbf{\Phi}_{t+\imath}^{(g-1)}$ and $\mathbf{\Phi}_{t+\imath,t+\imath+k}^{(g)} = \mathbf{\Phi}_{t+\imath,t+\imath+k}^{(g-1)}$ for all i and k. Then for $\imath = 0$, the initial state covariance matrix of cohort g, $\mathbf{\Phi}_{t_1}^{(g)}$, can be written as

$$\begin{aligned} \mathbf{\Phi}_{t_1}^{(g)} &= \mathbf{\Phi}_{t_1}^{(g-1)} \\ &= \mathcal{A}_{t_1,t_0}^{(g-1)}\mathbf{\Phi}_{t_0}^{(g-1)}\mathcal{A}_{t_1,t_0}^{(g-1)\prime} + \sum_{k=t_0}^{t-1}\mathcal{A}_{t_1,k+1}^{(g-1)}\mathbf{Q}_k^{(g-1)}\mathcal{A}_{t_1,k+1}^{(g-1)\prime} \\ &\equiv \mathcal{V}^{(g,g-1)}\ , \end{aligned} \tag{4.31}$$

which for $\imath = 1$ results in,

$$\begin{aligned} \mathbf{\Phi}_{t_1+1}^{(g)} &= \mathbf{\Phi}_{t_1+1}^{(g-1)} = \\ \mathbf{A}_{t_1}^{(g)}\mathcal{V}^{(g,g-1)}\mathbf{A}_{t_1}^{(g)\prime} + \mathbf{Q}_{t_1}^{(g)} &= \mathbf{A}_{t_1}^{(g-1)}\mathcal{V}^{(g,g-1)}\mathbf{A}_{t_1}^{(g-1)\prime} + \mathbf{Q}_{t_1}^{(g-1)}\ . \end{aligned} \tag{4.32}$$

It shows the recursive relation over time of the latent covariance matrices, again involving the state transition matrix. Here also the initialization problem, which concerns the estimation of the initial latent covariance matrix $\mathbf{\Phi}_{t_1}^{(g)}$ (see Equation 4.31), not having any past information in cohort g, is solved by handing over this information from cohort $g-1$ in $\mathcal{V}^{(g,g-1)}$ to cohort g. Note that the constraints between latent off-diagonal covariance matrices $\mathbf{\Phi}_{t+\imath,t+\imath+k}^{(g)} = \mathbf{\Phi}_{t+\imath,t+\imath+k}^{(g-1)}$, k the between time points interval with $1 \leq k \leq max(i)$, imply that $\mathbf{A}_{t_1+\imath}^{(g)} = \mathbf{A}_{t_1+\imath}^{(g-1)}$. For $i = 0$ and $k = 1$, for example,

$$\begin{aligned} \mathbf{\Phi}_{t_1,t_1+1}^{(g)} &= \mathbf{\Phi}_{t_1,t_1+1}^{(g-1)} = \\ \mathbf{\Phi}_{t_1}^{(g)}\mathbf{A}_{t_1}^{(g)\prime} &= \mathbf{\Phi}_{t_1}^{(g-1)}\mathbf{A}_{t_1}^{(g-1)\prime} \ , \end{aligned} \tag{4.33}$$

also implying that $\mathbf{Q}_{t_1+\imath}^{(g)} = \mathbf{Q}_{t_1+\imath}^{(g-1)}$ (see Equation 4.32). The constraints formula for restricting the latent variances and covariances between two adjacent cohorts is (cfr. Equation 4.16)

$$\mathbf{S}_1^{(g)}\mathbf{\Phi}_{\mathbf{x}}^{(g)}\mathbf{S}_1^{(g)\prime} - \mathbf{S}_2^{(g-1)}\mathbf{\Phi}_{\mathbf{x}}^{(g-1)}\mathbf{S}_2^{(g-1)\prime} = \mathbf{0} \ , \tag{4.34}$$

with $\mathbf{S}_1^{(g)}$ and $\mathbf{S}_2^{(g-1)}$ selecting the appropriate elements to be constrained. $\mathbf{S}_1^{(g)}$ and $\mathbf{S}_2^{(g-1)}$ have the same form as $\mathbf{S}_1$ and $\mathbf{S}_2$ above, but with roworders equal to $mT_{g,g-1}$.

Latent means and variances-covariances can both be constrained to be common for two adjacent cohorts by combining constraints formulas 4.30 and 4.34 in one analysis. Because $\mathbf{A}_{t_1}^{(g)} = \mathbf{A}_{t_1}^{(g-1)}$, it is then implied that $\mathbf{b}_{t_1}^{(g)} = \mathbf{b}_{t_1}^{(g-1)}$ (see Equation 4.29).

Constraints concerning common cohort characteristics may be combined with within-cohort stationarity constraints. For example, combining constraints formulas 4.15 and 4.30, a first-order nonzero-means stationary OCD results, yielding equal latent means within and between cohorts, while still allowing the autoregressive and cross-lagged parameters and latent intercepts to be time-varying. Restricting then between cohorts $\mathbf{A}_{t_1+\imath}^{(g)} = \mathbf{A}_{t_1+\imath}^{(g-1)}$ yields $\mathbf{b}_{t_1+\imath}^{(g)} = \mathbf{b}_{t_1+\imath}^{(g-1)}$, and in case of common latent covariance matrices (constraints formula 4.34) $\mathbf{Q}_{t_1+\imath}^{(g)} = \mathbf{Q}_{t_1+\imath}^{(g-1)}$. Additionally constraining the model parameters to be time-invariant, one writes $\mathbf{A}^{(g)} = \mathbf{A}^{(g-1)}$, $\mathbf{b}^{(g)} = \mathbf{b}^{(g-1)}$ and $\mathbf{Q}^{(g)} = \mathbf{Q}^{(g-1)}$ for all cohorts g. Then, if the model of the first cohort is second-order stationary by the additional specification of the initializing condition in Equation 4.31, the whole model over all cohorts becomes second-order stationary.

4.6 Examples

To illustrate the application of nonstandard constraints in SEM state space modeling, two data sets were simulated, assuming a second-order zero-means stationary process, and a second-order nonzero-means stationary process. Each data set comprised two different age cohorts. For implementation of the constraints the SEM program Mx was used. In one cohort (cohort 1) five analyses were performed, involving, respectively, a zero-means process (implying first-order stationarity) and a nonzero-means (nonstationary) process, a first-order nonzero-means stationary process, and a second-order zero- and nonzero-means stationary process. For the OCD, consisting of cohorts 1 and 2, four analyses were performed for the zero-means processes, and six analyses for the nonzero-means processes. With regard to these processes, additional between cohort constraints were imposed upon the model, starting with a model 1) not involving any between cohort constraints, and subsequently specifying 2) common measurement characteristics, 3) common latent means in case of nonzero-means, 4) common latent covariance matrices, 5) within-cohort first-order stationarity in case of nonzero-means, and finally, 6) within-cohort second-order stationarity.

The data were generated on the basis of simulated latent variable processes. First, the model equations, the initial means and covariance matrices, and the parameter values were specified (see Table 4.1). Then a number of iterations (No. it. in Table 4.1) or time propagation steps were performed until the processes were stationary up to order two, meaning that the latent covariance and moment matrices had a *Toeplitz* structure. The observed covariance and moment matrices were obtained by applying the parameter matrices of the measurement models, and were then used to randomly generate data for a sample size of $N = 500$. Four raw data sets were generated for each of the four columns in Table 4.1. Table 4.1 only displays the first columns of the latent covariance and moment matrices because these contain all the information of a *Toeplitz* structured matrix.

Figure 4.2 displays the nonzero-means model of cohort 1. The squares represent the observed variables at different points in time, being explained by the latent variables represented by the circles. $\sigma^{2(1)}_{x_{t_0}}$ represents the initial latent variance with $\mu^{(1)}_{x_{t_0}}$ the initial latent mean. $q^{(1)}_{t_0}$, $q^{(1)}_{t_0+1}$ and $q^{(1)}_{t_0+2}$ represent the process error variances at successive points in time, and $a^{(1)}_{t_0}$, $a^{(1)}_{t_0+1}$ and $a^{(1)}_{t_0+2}$ the between time points autoregressive coefficients. The latent intercepts are given by $b^{(1)}_{t_0}$, $b^{(1)}_{t_0+1}$, and $b^{(1)}_{t_0+2}$. The factorloadings c_1, c_2, measurement origins d_1, d_2, and measurement error variances r_1, r_2 were assumed to be equal for the same tests used over time. For identification the d_1 parameter in the nonzero-means model was fixed to zero. Combined with the unit factorloading c_1, it made the initial latent mean $\mu^{(1)}_{x_{t_0}}$ equal the initial observed mean $\mu^{(1)}_{y_1,t_0}$, and the initial latent variance equal the true initial observed variance $\sigma^{2(1)}_{x_{t_0}} = \sigma^{2(1)}_{y_1,t_0} - r_1$.

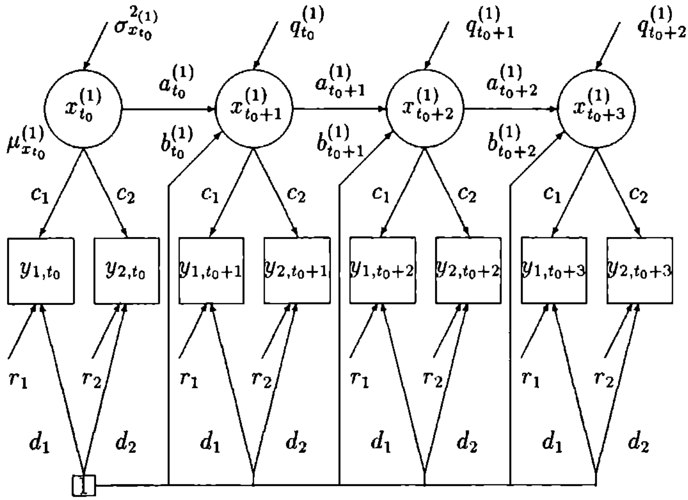
In the nonzero-means model a total of 16 parameters had to be estimated on the basis of 9 observed variables (the unit input-variable included), leaving 29 degrees of freedom. In the zero-means model the mean and intercept parameters were omitted from the model, as well as the moment of the unit input-variable, and d_2. A total of 10 parameters had to be estimated on the basis of 8 observed variables, leaving 26 degrees of freedom.

Tab. 4.1: True parameter values and latent stationary covariance and moment matrices of the zero- and nonzero-means model of cohort 1 and 2 ($i \geq 0$, and No. it. the number of iterations needed to reach stationarity).

	Zero-means		Nonzero-means	
	Cohort 1	Cohort 2	Cohort 1	Cohort 2
a_{t_0+i}	.9	.9	.9	.9
b_{t_0-1}	0	0	30	32
b_{t_0+i}	0	0	4	4.2
q_{t_0+i}	20	19	20	19
$\sigma^2_{x_{t_0}}$	200	199	200	199
c_1	1	1	1	1
c_2	.7	.7	.8	.8
d_1	0	0	0	0
d_2	0	0	-2	-2
r_1	10	12	10	12
r_2	8	9	8	9
Latent variances	105.26316	100.00		
Latent covariances	94.736842	90.00		
	85.263158	81.00		
	76.736842	72.90		
	No. it. 45	No. it. 44		
Latent moments			1705.2632	1864.00
			1694.7368	1854.00
			1685.2632	1845.00
			1676.7368	1836.90
Latent means			40.0	42.0
			No. it. 84	No. it. 83

Table 4.2 shows the parameter estimates and model fit of both the zero- and nonzero-means models. Evidently, these models fit very well. Almost all estimated values are within the intervals of true values (see Table 4.1) ± 1 standard error. The upper section of Table 4.3 contains the results of the other SEM models of cohort 1. Note that the first-order zero-means stationary model is the same as the zero-means model of Table 4.2. The second-order zero-means stationary model has five additional degrees of freedom in comparison to the corresponding first-order stationary model, because of the time-invariance of the autoregressive coefficients and process error variances implied by constraints formula 4.16, and because of

Fig. 4.2: Nonzero-means SEM model of cohort 1.



one extra degree of freedom resulting from the restrictions on the latent variances and covariances (see Equation 4.24). The χ^2-difference test (Jöreskog & Sörbom, 1989, p. 244-246): $\chi^2 = \chi^2_{31} - \chi^2_{26} = 3.02$ which for *df* $= 5$ is not significant, indicates that the second-order stationary zero-means model better fits the data. This is in accordance with the true parameter values in Table 4.1.

The first-order nonzero-means stationary model of cohort 1 (Table 4.3) has $T-1 = 3$ additional degrees of freedom in comparison to the nonzero-means model of Table 4.2, each of the restrictions, $E(x_{t_0}) = E(x_{t_0+1})$, $E(x_{t_0+1}) = E(x_{t_0+2})$, $E(x_{t_0+2}) = E(x_{t_0+3})$ yielding one degree of freedom. With a χ^2-value of 32.33 and *df* $= 32$ the model fits reasonably well. In applying constraints formula 4.16 a second-order stationary process is estimated. Time-invariance is implied for three autoregressive coefficients, three process error variances, and three latent intercepts, yielding six additional degrees of freedom. Furthermore, as Equation 4.24 indicates, one degree of freedom results from the restrictions on the initial latent variances and covariances. On the basis of $\chi^2 = \chi^2_{39} - \chi^2_{32} = 15.17$ which for *df* $= 7$ is not significant, the second-order nonzero-means stationary model is not rejected as it should not (see true values in Table 4.1).

Next the OCD was estimated and tested for the zero- and nonzero-means processes (lower section Table 4.3). Interest was in commonness of the two partially overlapping age-cohorts. Figure 4.3 displays the zero-means OCD. (time subscripts of y_1 and y_2 are omitted). The cohorts overlap at two points in time. That is, at $t_0 + 2$ and $t_0 + 3$, the ages of the cohorts coincide. Whereas each cohort covers four measurement times, the cohorts together cover a total of six measurement times. Only four measurement time points are needed to provide the information of six points in time.

Tab. 4.2: Parameter estimates of the zero-means and nonzero-means model of cohort 1 (N = 500), with standard errors between parentheses.

Latent autoregressive parameters		a_{t_0}	a_{t_0+1}	a_{t_0+2}
Zero-means		.886	.926	.873
		(.026)	(.026)	(.026)
Nonzero-means		.852	.885	.845
		(.025)	(.024)	(.024)
Initial latent variance and unexplained variances	$\sigma^2_{x_{t_0}}$	q_{t_0}	q_{t_0+1}	q_{t_0+2}
Zero-means	97.94	20.05	18.46	21.47
	(6.66)	(1.96)	(1.84)	(2.05)
Nonzero-means	113.62	23.46	19.02	17.97
	(7.61)	(2.05)	(1.75)	(1.70)
Factor loadings	c_1	c_2		
Zero-means	1.000	.691		
		(.009)		
Nonzero-means	1.000	.790		
		(.009)		
Measurement error variances	r_1	r_2		
Zero-means	9.54	8.22		
	(.67)	(.39)		
Nonzero-means	9.22	7.55		
	(.61)	(.41)		
Initial latent mean and intercepts	$\mu_{x_{t_0}}$	b_{t_0}	b_{t_0+1}	b_{t_0+2}
Nonzero-means	40.57	5.68	4.28	6.44
	(.49)	(1.03)	(.98)	(.97)
Origins observed variables	d_1	d_2		
Nonzero-means	0.000	-1.601		
		(.365)		
	χ^2	*df*	*AIC*	
Zero-means	24.20	26	-27.80	
Nonzero-means	28.03	29	-29.98	

Starting with the zero-means OCD, a multi-sample SEM analysis was performed without the application of any between- and within-cohort constraints. It yielded a reasonably well fitting model, χ^2 = 56.27 and *df* = 52 (Table 4.3).

Next, several tests of commonness in the two cohorts were performed. As the data sets were generated as samples from 'very' similar true processes (see Table 4.1), no or only very few rejections of commonness were expected for sample sizes of $N = 500$. First, commonness in measurement scale characteristics was tested at the overlapping time points: $\overline{\mathbf{C}}^{(1)} = \overline{\mathbf{C}}^{(2)}$. A total of 19 parameters had to be estimated leaving 53 degrees of freedom. The $\chi^2 = \chi^2_{53} - \chi^2_{52} = .73$ for $df = 1$ is not significant. The AIC values (Akaike, 1974) indicate a slightly better fit. The commonness in measurement characteristics is in accordance with the true values in Table 4.1.

Tab. 4.3: Model fit of the first- and second order zero-means and nonzero-means stationary model for cohort 1 and for the OCD, ($N = 500$).

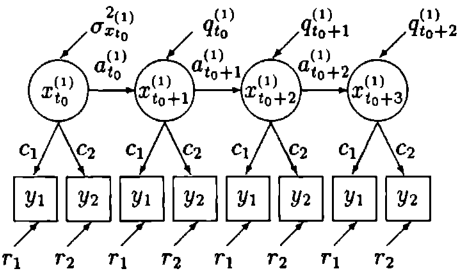
SEM	Zero-means			Nonzero-means			
	χ^2	df	AIC	χ^2	df	AIC	
Cohort 1	24.20	26	-27.80	32.33	32	-31.67	1^{st} order stationary
	27.22	31	-34.78	47.50	39	-30.50	2^{nd} order stationary
OCD	Zero-means			Nonzero-means			
	χ^2	df	AIC	χ^2	df	AIC	
	56.27	52	-47.73	52.54	58	-63.46	No constraints
	57.00	53	-49.00	52.85	60	-67.15	$\overline{\mathbf{C}}^{(1)} = \overline{\mathbf{C}}^{(2)}$, $\overline{\mathbf{D}}^{(1)} = \overline{\mathbf{D}}^{(2)}$
				68.06	62	-55.94	Common latent means
	59.53	56	-52.47	70.12	66	-61.88	Common latent variances-covariances
				73.18	71	-68.82	1^{st}order stationary
	71.47	65	-58.53	94.61	80	-65.39	2^{nd} order stationary

Next, commonness in the latent variances and covariances at the overlapping time points was tested for, leading to three additional degrees of freedom: that is, $a^{(1)}_{t_0+2} = a^{(2)}_{t_0+2}$ and $q^{(1)}_{t_0+2} = q^{(2)}_{t_0+2}$, and $\varphi^{(1)}_{t_0+2} = \varphi^{(2)}_{t_0+2}$ implying also that $\varphi^{(1)}_{t_0+3} = \varphi^{(2)}_{t_0+3}$ (cfr. Equations 4.31 and 4.32). It gave a $\chi^2 = \chi^2_{56} - \chi^2_{53} = 2.53$ which is not significant for $df = 3$ (see Table 4.3).

In a final analysis, it was tested whether the processes were additionally second-order stationary implying that, $a^{(1)}_t = a^{(2)}_t = a$ and $q^{(1)}_t = q^{(2)}_t = q$ for all $t \geq t_0$, yielding eight additional degrees of freedom, and one extra degree of freedom because of the conditions on the initial latent variances and covariances of the first cohort. The results gave a $\chi^2 = \chi^2_{65} - \chi^2_{56} = 11.94$, which again is not significant for $df = 9$. In fact, the final model fitted quite well, also indicated by the AIC value of -58.53, being smaller than the AIC value of the previous model.

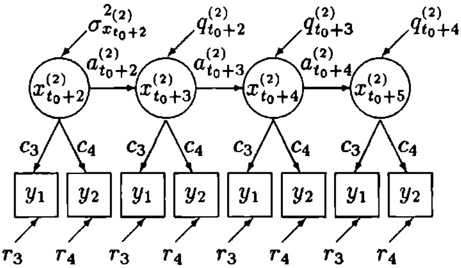
The analysis proceeded with the nonzero-means OCD. The model differs from

Fig. 4.3: Zero-means overlapping cohort model, cohort 1 and 2.



Cohort 1

Cohort 2



the OCD in Figure 4.3 by the addition of the unit input-variable (see Figure 4.2). Again the analyses started with the nonconstrained OCD. First, commonness in the measurement scale characteristics was tested for; $\overline{\mathbf{C}}^{(1)} = \overline{\mathbf{C}}^{(2)}$ and $\overline{\mathbf{D}}^{(1)} = \overline{\mathbf{D}}^{(2)}$. It gave a nonsignificant increase of $\chi^2 = \chi^2_{60} - \chi^2_{58} = .31$ for $df = 2$ (see Table 4.3).

The model was further analyzed by additionally testing whether the latent means coincided at the overlapping time points. Because of the additional constraints $E(x_{t_0+2})^{(1)} = E(x_{t_0+2})^{(2)}$ and $E(x_{t_0+3})^{(1)} = E(x_{t_0+2})^{(3)}$, it led to two additional degrees of freedom. With a χ^2-value of 68.06 and $df = 62$, the model still fitted reasonably well.

Next, commonness in the variances and covariances at the overlapping time points was additionally tested for. It yielded four additional degrees of freedom, because of $a^{(1)}_{t_0+2} = a^{(2)}_{t_0+2}$, $q^{(1)}_{t_0+2} = q^{(2)}_{t_0+2}$, and $b^{(1)}_{t_0+2} = b^{(2)}_{t_0+2}$, and because of $\varphi^{(1)}_{t_0+2} = \varphi^{(2)}_{t_0+2}$ implying that $\varphi^{(1)}_{t_0+3} = \varphi^{(2)}_{t_0+3}$ (cfr. Equations 4.31 and 4.32). With a $\chi^2 = \chi^2_{66} - \chi^2_{62} = 2.06$, which is not significant for $df = 4$, and an AIC-decrease of 5.94, a slightly better fitting OCD resulted.

Two final analyses were performed, involving a number of highly complex constraints. First, first-order stationarity over both cohorts was additionally tested for. It yielded $2 \times (T-1) - 1 = 5$ extra degrees of freedom. That is, because of the common latent means $E(x_{t_0+2})^{(1)} = E(x_{t_0+2})^{(2)}$ and $E(x_{t_0+3})^{(1)} = E(x_{t_0+3})^{(2)}$

implemented in a previous step, $E(x_{t_0+2})^{(2)} = E(x_{t_0+3})^{(2)}$ is already implied by $E(x_{t_0+2})^{(1)} = E(x_{t_0+3})^{(1)}$. A nonsignificant $\chi^2 = \chi^2_{71} - \chi^2_{66} = 3.06$ for $df = 5$ resulted. Second, second-order stationarity constraints were finally applied. Because of $a_t^{(g)} = a$ and $q_t^{(g)} = q$, it yielded 8 additional degrees of freedom plus one extra because of constraining the latent variances and covariances over time (see Equation 4.24). With a $\chi^2 = \chi^2_{80} - \chi^2_{71} = 21.43$ for $df = 9$, which is not significant again, a slightly worse fit was obtained. The AIC value increased from -68.82 to -65.39. To sum up, the results in Table 4.3 show that the hypothesis of commonness in the latent parts of the OCDs and stationarity up to order two can be confirmed. In testing for cohort commonness, the small differences in the true values of Table 4.1 are not significant in the corresponding random samples.

However, as can be seen from Table 4.4, with the results for a sample size of $N = 5000$, these differences tend to become significant. Especially the differences in the latent nonzero-means of the two overlapping cohorts are salient in Table 4.4. Additional constraints for the nonzero-means model, that is, common latent variances-covariances and first-order stationarity, make that the increase in the χ^2-value is higher than the increase in the number of degrees of freedom.

Tab. 4.4: Model fit of the first- and second order zero-means and nonzero-means stationary model for cohort 1 and for the OCD, ($N = 5000$).

SEM	Zero-means			Nonzero-means			
	χ^2	df	AIC	χ^2	df	AIC	
Cohort 1	14.20	26	-37.80	21.31	32	-42.69	1^{st} order stationary
	21.35	31	-40.65	24.08	39	-53.92	2^{nd} order stationary
OCD	Zero-means			Nonzero-means			
	χ^2	df	AIC	χ^2	df	AIC	
	54.64	52	-49.36	44.88	58	-71.12	No constraints
	55.05	53	-50.95	46.72	60	-73.28	$\overline{C}^{(1)} = \overline{C}^{(2)}$, $\overline{D}^{(1)} = \overline{D}^{(2)}$
				114.76	62	-9.24	Common latent means
	58.72	56	-53.28	124.55	66	-7.46	Common latent variances-covariances
				143.12	71	1.12	1^{st} order stationary
	70.90	65	-59.10	145.60	80	-14.41	2^{nd} order stationary

4.7 Conclusion

It is shown how nonstandard linear and nonlinear constraints are applied in two central areas of SEM-SSM modeling of panel data. The testing of both stationarity and of overlapping cohort commonness in the OCD are made possible by the availability of nonstandard constraints in recent SEM programs. Only some SEM software packages as, for example, Mx do allow complex nonstandard constraints to be made by means of matrix algebraic expressions. As applications of (longitudinal) SEM models are becoming more comprehensive, there will be a greater need for SEM programs which can handle these type of problems.

The OCD shortens the data collection period and is much less vulnerable for panel attrition. These advantages of the OCD, however, are based on cohort commonness. Thus not only the testing and estimation of common model parameters, as is traditionally possible in SEM by means of standard constraints, but also simultaneously of common model implied latent means and variances-covariances is essential. The nonlinearities have been shown to reside in the way the initial latent means and variances-covariances in subsequent cohorts are connected to the parameters in preceding cohorts.

Chapter 5

Filtering and smoothing on the basis of the SEM state space model [1]

Abstract

The basic discrete-time SSM is introduced and shown to represent a very general class of dynamic models. For the extended (additional input-effects) version of the SSM it is shown in detail how it can be represented as a SEM model and estimated by means of a SEM program. The essentials of the discrete-time Kalman filter in comparison to the traditional cross-sectional factor score estimators are explained, stressing unbiasedness considerations and initialization of the Kalman filter. The Kalman smoother is discussed next, especially as regards its connection to the 'overall' regression estimator and unbiasedness as well. Finally, two examples are presented which show how the Kalman filter and Kalman smoother can be further enhanced for behavioral science applications. The first example shows how the traditional 'zero means' SEM model is replaced by the more complicated 'structured means' SEM model, enabling the Kalman filter or Kalman smoother to estimate absolute as well as relative developmental curves. The second example shows how random subject effects or trait variables, defining over time constant subject specific intercept values, are included in the SEM model and can be estimated by means of the Kalman filter.

5.1 Introduction

Modeling efforts in the behavioral sciences, especially on the basis of SEM, are typically theoretically orientated by aiming at causal interpretations of reality. After

[1] Adapted version of Oud, J H L , van Leeuwe, J F J , & Jansen, R A R G (1993) Kalman filtering in discrete and continuous time based on longitudinal LISREL models In J H L Oud & A W van Blokland-Vogelesang (Eds), *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp 3-26) Nijmegen ITS

drawing the causal picture in accordance with or inspired by the model and giving recommendations for the construction of improved models, most studies finish. When in longitudinal research the model is written in the form of the dynamic SSM, it can be made practically useful by employing the Kalman filter (Kalman, 1960; Kalman & Bucy, 1961) and Kalman smoother (Lewis, 1986; Rauch, Tung & Striebel, 1965) for the optimal estimation of individual latent developmental curves.

While the Kalman filter, which is an on-line estimator, utilizes past and current information, the Kalman smoother additionally utilizes future information for the estimation of latent developmental values at previous points in time. Because of this, filtering is sometimes preferred to smoothing. In case observations are lacking, the filter provides optimal latent estimates by its built-in model based predictor. In addition, by comparing the estimated actual development under some intervention condition (intervention curve) with the previously model based prediction of development (control curve), the evaluation of intervention effects in individual cases becomes possible. In this respect the Kalman filter is more appropriate then the Kalman smoother: the smoother state estimates at a previous time point change as a result of the intervention at later points in time (more future information becomes available), and thus the latent state estimates are influenced by the intervention even before it takes place. In case of missing values, however, the Kalman smoother is to be preferred as it utilizes all available information for the optimal reconstruction of the latent score lacking current observations (e.g. Jansen & Oud, 1995; Oud & Jansen, 1996).

In earlier studies (Molenaar & Oud, 1991; Oud, Van den Bercken, & Essers, 1990) the advantages of the Kalman filter as a factor score estimator were discussed. Also, the SSM, originating from control theory, was shown to be translatable into SEM form (Jöreskog & Sörbom, 1989). In this way the SSM parameters, entering the Kalman filter, were shown to be estimable for behavioral science data by means of a SEM program. The Kalman filter was implemented in an early version of the pupil monitoring system LISKAL (Oud, Mommers, & Heijmans, 1991), which was based on a longitudinal SEM model for reading and spelling. Replacement of the cross-sectional Bartlett estimator (Lawley & Maxwell, 1971, p. 110) by the Kalman filter considerably reduced the estimation error variances of the latent reading and spelling scores. In a recent version of LISKAL the filter is applied on the basis of longitudinal SEM modeling of decoding speed, reading comprehension, spelling, arithmetic, and vocabulary (Aarnoutse, van Leeuwe, Oud, Voeten & van Kan, 1996a; Aarnoutse, van Leeuwe, Voeten, van Kan & Oud, 1996b).

In this article the discrete-time linear SSM and the SEM model are explained first. The SSM is shown to represent a very general class of dynamic models. Autoregressive moving-average (ARMA) models of arbitrary order can be reformulated as to fit into the SSM model. The SSM model allows the inclusion of random subject-effect or trait variables, also known as 'unobserved heterogeneity'. Finally, so-called (fixed) input-variables can be specified, allowing many different types of inputs to become part of the model. A special and relatively simple case is the inclusion of the unit input-variable, to be employed for the modeling of means processes. This case corresponds to the structured means SEM model. Second, the essentials of the Kalman filter in comparison to the regression and Bartlett factor score estimators are explained, stressing unbiasedness considerations. Third, the Kalman smoother is discussed as well as how it relates to the so-called 'overall' regression estimator and to one form of unbiasedness. Finally, two examples are presented which show how the usefulness of the Kalman filter and smoother can be further enhanced for behavioral science applications. The first example discusses Kalman filtering on the basis of the structured means SEM model, which enables the estimation of absolute as well as relative developmental curves. The SEM model construction and identification of the means parameters are illustrated. The second example discusses filtering and smoothing on the basis of the random subject effects or 'state-trait' model. In adding trait variables to the SSM, traits or constant subject specific intercept values can be estimated by application of the Kalman filter. The state-trait model leads to each subject having its own intercept value and, therefore, its own mean curve towards which its state regresses or from which it egresses. It can be statistically tested whether traits underlie development and must be included in the model.

5.2 State space modeling

5.2.1 Basic model

The discrete-time linear stochastic SSM consists of two equations: the measurement equation (Equation 5.2), which is equivalent to the factor model equation of factor analysis with $\mathbf{y}_t$ the vector of observed variables and $\mathbf{C}_t$ the factor pattern matrix, and the state equation (Equation 5.1), which describes the dependency of the latent state variables or factors in vector $\mathbf{x}_t$ on their lagged values in vector $\mathbf{x}_{t-1}$:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{w}_{t-1} \qquad \text{with} \qquad cov(\mathbf{w}_{t-1}) = \mathbf{Q}_{t-1} \ , \qquad (5.1) \\ \mathbf{y}_t &= \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t \qquad \text{with} \qquad cov(\mathbf{v}_t) = \mathbf{R}_t \ . \qquad (5.2) \end{aligned}$$

Matrix $\mathbf{A}_{t-1}$ contains the autoregressive and cross-lagged effects between the state variables at successive discrete time points t and $t-1$; $t, t-1 \in \{t_0, t_0+1, \ldots, t_0+T-1\}$ for integers t_0 and $T \geq 2$ with t_0 the initial time point and T the total number of time points considered. Because of the time subscripts of the four model matrices the model is specified time-varying. This means that the causal characteristics of the system are allowed to change over time. The following assumptions are made concerning the process errors in successive vectors $\mathbf{w}_{t-1}$ and the measurement errors in successive vectors $\mathbf{v}_t$: a) zero expectations, b) zero covariances between vectors (covariances within vectors are in matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$), c) zero covariances with the initial state $\mathbf{x}_{t_0}$. Further, d) the initial state is supposed to have zero expectations, and finally, e) the random variables in the error vectors and the initial state are supposed to be jointly multinormally distributed.

Because the complete dynamics of the evolution of factor scores over time in addition to their instantaneous manifestation in the observables $\mathbf{y}_t$ (called output variables in the state space approach) is recursively described in as few as four model matrices, the SSM represents a very parsimoniously formulated though general dynamic factor model. When instead of the state equation the more complicated, so-called structural equation is formulated for the evolution of the factor scores or state variables over time,

$$\mathbf{x}_t = \mathbf{K}_t\mathbf{x}_t + \mathbf{A}^{\circ}_{t-1}\mathbf{x}_{t-1} + \mathbf{w}^{\circ}_{t-1} , \tag{5.3}$$

the state equation can nevertheless be derived as its reduced form. Premultiplying both sides by $\mathbf{M}_t = (\mathbf{I} - \mathbf{K}_t)^{-1}$ where $\mathbf{I} - \mathbf{K}_t$ is assumed to be nonsingular, Equation 5.3 reduces to Equation 5.1: $\mathbf{A}_{t-1} = \mathbf{M}_t\mathbf{A}^{\circ}_{t-1}$, $\mathbf{w}_{t-1} = \mathbf{M}_t\mathbf{w}^{\circ}_{t-1}$, $\mathbf{Q}_{t-1} = \mathbf{M}_t\mathbf{Q}^{\circ}_{t-1}\mathbf{M}'_t$. Depending on whether or not the matrix $\mathbf{K}_t$, containing instantaneous effects between the state variables, can be chosen as a strictly triangular matrix in combination with a diagonal matrix $\mathbf{Q}^{\circ}_{t-1}$ (covariance matrix of $\mathbf{w}^{\circ}_{t-1}$), the structural equation model is called recursive or interdependent.

Although the causal interpretation of the instantaneous effects in $\mathbf{K}_t$ of Equation 5.3 poses some difficulties, especially in interdependent systems where the causal chain principle does not hold (Wold, 1954), their specification in discrete-time models is a natural consequence of the choice of a discrete time interval between observations that overlaps the minimum time lag in causal dependencies (Bergstrom, 1984, p. 1147). In estimating the SSM, one could skip the structural equation and directly address Equation 5.1. However, estimating first the structural form matrices in Equation 5.3 with the appropriate theoretical restrictions incorporated and then deriving the reduced form matrices, typically leads to a considerable precision gain over estimating directly the matrices in Equation 5.1 but unconstrained (Bergstrom, 1984, pp. 1146-1147).

5.2.2 ARMA extension

The whole class of observed as well as latent ARMA models with arbitrary maximum lag $k > 1$ is covered by the SSM. An example is the following latent ARMA model (latent because of the specification $\mathbf{C}_t \neq \mathbf{I}$ and $\mathbf{v}_t \neq \mathbf{0}$) with maximum lag $k = 2$ (2nd-order autoregressive because of $\mathbf{A}^{\bullet}_{t-2}$, and 1st-order moving average because of $\mathbf{G}^\bullet_{t-2}$):

$$\mathbf{x}^\bullet_t = \mathbf{A}^\bullet_{t-1}\mathbf{x}^\bullet_{t-1} + \mathbf{A}^\bullet_{t-2}\mathbf{x}^\bullet_{t-2} + \mathbf{G}^\bullet_{t-1}\mathbf{w}^\bullet_{t-1} + \mathbf{G}^\bullet_{t-2}\mathbf{w}^\bullet_{t-2}\ , \tag{5.4}$$

$$\mathbf{y}_t = \mathbf{C}^\bullet_t\mathbf{x}^\bullet_t + \mathbf{v}_t\ . \tag{5.5}$$

This model is reformulated in correct state space form as follows:

$$\underbrace{\begin{bmatrix} \mathbf{x}^\bullet_t \\ \mathbf{w}^\bullet_t \\ \mathbf{x}^\bullet_{t-1} \\ \mathbf{w}^\bullet_{t-1} \end{bmatrix}}_{\mathbf{x}_t} = \underbrace{\begin{bmatrix} \mathbf{A}^\bullet_{t-1} & \mathbf{G}^\bullet_{t-1} & \mathbf{A}^\bullet_{t-2} & \mathbf{G}^\bullet_{t-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{A}_{t-1}} \underbrace{\begin{bmatrix} \mathbf{x}^\bullet_{t-1} \\ \mathbf{w}^\bullet_{t-1} \\ \mathbf{x}^\bullet_{t-2} \\ \mathbf{w}^\bullet_{t-2} \end{bmatrix}}_{\mathbf{x}_{t-1}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{w}^\bullet_t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{w}_{t-1}}\ , \tag{5.6}$$

$$\begin{aligned} \mathbf{y}_t &= \begin{bmatrix} \mathbf{C}^\bullet_t & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{x}_t + \mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{C}_t\ \mathbf{x}_t + \mathbf{v}_t \end{aligned} \tag{5.7}$$

The idea behind the reformulation is that by putting the lag 1 vectors $\mathbf{x}^\bullet_{t-1}$ and $\mathbf{w}^\bullet_{t-1}$ in the newly defined current state $\mathbf{x}_t$ the lag 2 vectors $\mathbf{x}^\bullet_{t-2}$ and $\mathbf{w}^\bullet_{t-2}$ become available in the new lagged state $\mathbf{x}_{t-1}$. It is easily generalized to ARMA models of arbitrary maximum lag. Equations 5.6 and 5.7 show that the ARMA model (Equations 5.4 and 5.5) can be written in the form of the SSM model of Equations 5.1 and 5.2. In fact, it can be proven that any time-varying or time-invariant ARMA structure can be reproduced as a SSM model (e.g. Caines, 1988, p. 111).

It is noticed that as a consequence of the identities introduced in the model above, the new process error vector $\mathbf{w}_{t-1}$ contains elements identically zero and consequently its associated covariance matrix $cov(\mathbf{w}_{t-1})$ contains diagonal zeroes. For identification of the moving-average parameters in $\mathbf{G}^\bullet_{t-1}$ and $\mathbf{G}^\bullet_{t-2}$ of Equation 5.6, instead of or in addition to restricting these parameters themselves (e.g. $\mathbf{G}^\bullet_{t-1}$ diagonal), the process errors in successive matrices $\mathbf{w}^\bullet_t$ are specified as standardized uncorrelated variables: $cov(\mathbf{w}^\bullet_t) = \mathbf{I}$. Parameters in $\mathbf{G}^\bullet_{t-1}$ and $\mathbf{G}^\bullet_{t-2}$ are then providing scaling factors accounting for arbitrary variances and possibly nonzero covariances. For identification and parameter estimation of ARMA models by means of SEM see Oud and Jansen (1995).

5.2.3 State-trait model

The SSM including trait variables is as follows,

$$\mathbf{x}^\circ_t = \mathbf{A}^\circ_{t-1}\mathbf{x}^\circ_{t-1} + \boldsymbol{\xi} + \mathbf{w}^\circ_{t-1}\ , \tag{5.8}$$

$$\mathbf{y}_t = \mathbf{C}^\circ_t\mathbf{x}^\circ_t + \mathbf{v}_t\ . \tag{5.9}$$

with $\boldsymbol{\xi}$ the vector of trait variables. A trait variable may be specified for each dimension of the latent state vector. It is characterized as a random but constant over time intercept term and zero mean normally distributed over the population of subjects. As trait variables represent latent variables also, $\boldsymbol{\xi}$ can be put into the state vector $\mathbf{x}_t$.

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_t^\circ \\ \boldsymbol{\xi} \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_{t-1}^\circ & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1}^\circ \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{t-1}^\circ \\ \mathbf{0} \end{bmatrix}, \\ \mathbf{x}_t &= \mathbf{A}_{t-1} \mathbf{x}_{t-1} + \mathbf{w}_{t-1} \end{aligned} \tag{5.10}$$

$$\begin{aligned} \mathbf{y}_t &= \begin{bmatrix} \mathbf{C}_t^\circ & \mathbf{0} \end{bmatrix} \mathbf{x}_t + \mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{C}_t \mathbf{x}_t + \mathbf{v}_t \end{aligned} \tag{5.11}$$

Evidently, the state-trait model is just a special case of the SSM of Equations 5.1-5.2. Specification of $\boldsymbol{\xi}$ is such that it does not affect the vector of observables $\mathbf{y}_t$ but influence the state variables only ('unobserved heterogeneity'). As in accordance with Equation 5.8 the initial state vector must be expressed as $\mathbf{x}_{t_0}^\circ = \mathbf{A}_{t_0-1}^\circ \mathbf{x}_{t_0-1}^\circ + \boldsymbol{\xi} + \mathbf{w}_{t_0-1}^\circ$, $\boldsymbol{\xi}$ should also be considered part of $\mathbf{x}_{t_0}^\circ$. It leads to the initial state covariance matrix becoming:

$$\boldsymbol{\Phi}_{t_0} = \begin{bmatrix} \boldsymbol{\Phi}_{x_{t_0}^\circ} & \boldsymbol{\Phi}_{x_{t_0}^\circ,\xi} \\ \boldsymbol{\Phi}_{\xi,x_{t_0}^\circ} & \boldsymbol{\Phi}_{\xi} \end{bmatrix}. \tag{5.12}$$

Significance tests on the existence of the random constant subject effects in $\boldsymbol{\xi}$ can be easily performed by testing the trait variances in $\boldsymbol{\Phi}_\xi$ and initial state-trait covariances in $\boldsymbol{\Phi}_{x_{t_0}^\circ,\xi}$ to differ from zero.

Because the trait influences the propagation of the state, it is interesting to see which value a subject's state regresses to in the state-trait model. Defining the state transition matrix as $\mathcal{A}_{t,t_0} = \prod_{k=1}^{t-t_0} \mathbf{A}_{t-k}$, which is also defined for $t = t_0$: $\mathcal{A}_{t_0,t_0} = \mathcal{A}_{t,t} \equiv \mathbf{I}$ (Luenberger, 1979), and assuming stabilizing feedback between $t-1$ and t (all eigenvalues of $\mathbf{A}_{t-1}$ within the unit circle of the complex plane), the subject's state regression,

$$E(\mathbf{x}_t|\mathbf{x}_{t_0}, \boldsymbol{\xi}) = \mathcal{A}_{t,t_0}\mathbf{x}_{t_0} + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi}\ , \tag{5.13}$$

is between $t-1$ and t towards the subject specific mean $\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi}$. Because the first term at the right-hand side of Equation 5.13 decreases between $t-1$ and t its Euclidian distance to zero, the mean value a subject's state regresses to (or egresses from in case of increasing Euclidian distance), becomes the zero-initial-state mean

$$E(\mathbf{x}_t|\mathbf{x}_{t_0} = \mathbf{0}, \boldsymbol{\xi}) = \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi} \ . \tag{5.14}$$

The subject specific mean keeps a subject specific distance $\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi}$ from the zero population mean. The state-trait model in a sense specifies a $N = 1$ model for each subject in the population of subjects separately, preventing the state to regress to (or egress from) the population mean value as would be the case in longitudinal SEM models without any traits specified.

5.3 SSM with input effects

The SSM of Equations 5.1 and 5.2 can be extended by the specification of deterministic input-variables. It is realized by adding a vector $\mathbf{u}_{t-1}$ of input-variables with input-effect matrix $\mathbf{B}_{t-1}$ to the state equation, and analogously by adding $\mathbf{u}_t$ with input-effect matrix $\mathbf{D}_t$ to the measurement equation.

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1} \ , && (5.15)\\ \mathbf{y}_t &= \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \ . && (5.16) \end{aligned}$$

Equations 5.15 and 5.16, in fact, represent a more general SSM model. The latent states in vector $\mathbf{x}_t$ are stochastically driven by its lagged state vector $\mathbf{x}_{t-1}$ and error vector $\mathbf{w}_{t-1}$, and deterministically by the inputs in $\mathbf{u}_{t-1}$. It should be noted that the models of, respectively, Equations 5.4-5.5 and 5.8-5.9, which have been shown to be special cases of the SSM of Equations 5.1 and 5.2, can also be adapted as to include input-variables.

The SSM with input-effects, $\mathbf{B}_{t-1}\mathbf{u}_{t-1} \neq \mathbf{0}$ and $\mathbf{D}_t\mathbf{u}_t \neq \mathbf{0}$, implies a nonzero means development, $E(\mathbf{x}_t) \neq \mathbf{0}$ and $E(\mathbf{y}_t) \neq \mathbf{0}$ for $t > t_0$, while $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$ and $E(\mathbf{y}_{t_0}) \neq \mathbf{0}$ are allowed also. The expectations of the latent and observed variables or mean trajectories are expressed as follows,

$$\begin{aligned} E(\mathbf{x}_t) &= \mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \ , && (5.17)\\ E(\mathbf{y}_t) &= \mathbf{C}_t\mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \mathbf{C}_t \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k + \mathbf{D}_t\mathbf{u}_t \ , && (5.18) \end{aligned}$$

Equations 5.13 and 5.14 should be replaced by 5.19 and 5.20, showing that the subject's state

$$E(\mathbf{x}_t|\mathbf{x}_{t_0}, \boldsymbol{\xi}) = \mathcal{A}_{t,t_0}\mathbf{x}_{t_0} + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi} + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \ , \tag{5.19}$$

regresses towards (egresses from) the subject specific zero-initial-mean

$$E(\mathbf{x}_t|\mathbf{x}_{t_0} = \mathbf{0}, \boldsymbol{\xi}) = \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi} + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \ , \tag{5.20}$$

which keeps a subject specific distance $\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\boldsymbol{\xi}$ from the corresponding nonzero means trajectory $\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k$ (cfr. Equations 5.13 and 5.14) in the subpopulation of subjects sharing the same input history.

The model of Equations 5.15 and 5.16 allows different types of input-effects. In one special case, which is treated in section 5.7.1, only a single so-called unit input-variable is specified ($u_t = 1$ for all t) which is constant over time points as well as over subjects in the sample (Jöreskog & Sörbom, 1989, pp. 273-275). Here the vectors $\mathbf{b}_{t-1}$ represent latent growth intercepts and the vectors $\mathbf{d}_t$ location parameters (origins) of the measurement instruments. The model implies a means process which is common to all subjects in the sample. In another special case the input-variables are all constant over time ($\mathbf{u}_t = \mathbf{u}_{t-k}$ for all t and $k > 0$) but, apart from the unit input-variable, varying over subjects. It corresponds to the longitudinal SEM model with background variables (e.g. gender). In the general case, for which the derivation and estimation of the SEM model is considered here, additional input-variables are specified that vary over time points as well as over subjects. The flexibility of the model implies that from a subpopulation of subjects sharing the same input history only one subject could be present in the sample.

By writing the expectations of the initial latent and observed variables as

$$E(\mathbf{x}_{t_0}) = \mathbf{B}_{t_0-1}\mathbf{u}_{t_0-1} \ , \tag{5.21}$$

$$E(\mathbf{y}_{t_0}) = \mathbf{C}_{t_0}E(\mathbf{x}_{t_0}) + \mathbf{D}_{t_0}\mathbf{u}_{t_0} \ , \tag{5.22}$$

the initial state mean $E(\mathbf{x}_{t_0})$ is modeled by means of an extra matrix $\mathbf{B}_{t_0-1}$, to be specified zero except, in the case of $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, for the elements corresponding to the unit input-variable in $\mathbf{u}_{t_0-1}$. The value and identifiability of $E(\mathbf{x}_{t_0})$ depend on the choice of $\mathbf{D}_{t_0}$ as well as of the factor loading matrix $\mathbf{C}_{t_0}$. The choice of the latter additionally determines the value and identifiability of the initial state covariance matrix $\mathbf{\Phi}_{t_0} = E([\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})][\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]')$. In fact, these choices determine the measurement scales (origins and measurement units) of the latent state variables. For example, by specifying values 0 and 1 on specific places of, respectively, $\mathbf{D}_{t_0}$ and $\mathbf{C}_{t_0}$, the latent measurement scales are chosen equal to those of specific observed variables in $\mathbf{y}_{t_0}$. Special identification techniques are needed, however, to guarantee that the latent measurement scales maintain the same origins and measurement units across the whole time range (see section 5.7.1).

5.4 SEM formulation and parameter estimation

In order to derive the SEM model of the extended SSM, in Equations 5.15 and 5.16, these are written in the following form:

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{t-1} & \mathbf{A}_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{w}_{t-1} \end{bmatrix}, \tag{5.23}$$

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}_t & \mathbf{C}_t \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_t \end{bmatrix}. \tag{5.24}$$

In collecting all input-variables in the input-vector $\mathbf{u}$ but specifying the constant input-variables (e.g. the unit input-variable) and other exactly linearly related input-variables only once in $\mathbf{u}$, and defining

$$\begin{array}{lll} \boldsymbol{\eta} = [\mathbf{u}'\ \mathbf{x}']' & \text{with} & \mathbf{x} = [\mathbf{x}'_{t_0}\mathbf{x}'_{t_0+1}\cdots\mathbf{x}'_{t_0+T-1}]', \\ \boldsymbol{\zeta} = [\mathbf{u}'\ \mathbf{w}']' & \text{with} & \mathbf{w} = [[\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]'\mathbf{w}'_{t_0}\cdots\mathbf{w}'_{t_0+T-2}]', \\ \mathbf{y} = [\mathbf{u}'\ \mathbf{y}'_0]' & \text{with} & \mathbf{y}_0 = [\mathbf{y}'_{t_0}\mathbf{y}'_{t_0+1}\cdots\mathbf{y}'_{t_0+T-1}]', \\ \boldsymbol{\varepsilon} = [\mathbf{0}'\ \mathbf{v}']' & \text{with} & \mathbf{v} = [\mathbf{v}'_{t_0}\mathbf{v}'_{t_0+1}\cdots\mathbf{v}'_{t_0+T-1}]', \end{array}$$

the SEM model is derived:

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & = & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \overline{\mathbf{B}} & \overline{\mathbf{A}} \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & + & \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} & , \\ \boldsymbol{\eta} & = & \boldsymbol{B} & \boldsymbol{\eta} & + & \boldsymbol{\zeta} & \end{array} \tag{5.25}$$

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} & = & \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \overline{\mathbf{D}} & \overline{\mathbf{C}} \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} & , \\ \mathbf{y} & = & \boldsymbol{\Lambda} & \boldsymbol{\eta} & + & \boldsymbol{\varepsilon} & \end{array} \tag{5.26}$$

where all parameter matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, $\mathbf{C}_t$, $\mathbf{D}_t$ are put on the appropriate places in $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, $\overline{\mathbf{C}}$, $\overline{\mathbf{D}}$, respectively. Notice that in $\mathbf{x}$ the initial state $\mathbf{x}_{t_0}$ gets zero rows in $\overline{\mathbf{A}}$ but $\mathbf{B}_{t_0-1}$ in $\overline{\mathbf{B}}$ for modeling its mean $E(\mathbf{x}_{t_0})$, which therefore is subtracted from $\mathbf{x}_{t_0}$ in $\mathbf{w}$. From Equations 5.25-5.26 one derives

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} & = & \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \overline{\mathbf{C}}(\mathbf{I}-\overline{\mathbf{A}})^{-1}\overline{\mathbf{B}}+\overline{\mathbf{D}} & \overline{\mathbf{C}}(\mathbf{I}-\overline{\mathbf{A}})^{-1} \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} & . \\ \mathbf{y} & = & \boldsymbol{\Lambda}(\mathbf{I}-\boldsymbol{B})^{-1} & \boldsymbol{\zeta} & + & \boldsymbol{\varepsilon} & \end{array} \tag{5.27}$$

Defining

$$\mathbf{D}_0 \equiv \overline{\mathbf{C}}(\mathbf{I}-\overline{\mathbf{A}})^{-1}\overline{\mathbf{B}}+\overline{\mathbf{D}} \quad \text{and} \quad \mathbf{C}_0 \equiv \overline{\mathbf{C}}(\mathbf{I}-\overline{\mathbf{A}})^{-1},$$

Equation 5.27 becomes

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}_0 & \mathbf{C}_0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} . \tag{5.28}$$

Writing the random vector $\mathbf{y}_0$ in terms of $\mathbf{D}_0$ and $\mathbf{C}_0$:

$$\mathbf{y}_0 = \mathbf{D}_0\mathbf{u} + \mathbf{C}_0\mathbf{w} + \mathbf{v} ,$$

its mean and covariance matrix are found to be:

$$\begin{aligned} \boldsymbol{\mu}_0 = E(\mathbf{y}_0) &= \mathbf{D}_0\mathbf{u} , && (5.29) \\ \boldsymbol{\Sigma}_0 = cov(\mathbf{y}_0) &= E[(\mathbf{y}_0 - \boldsymbol{\mu}_0)(\mathbf{y}_0 - \boldsymbol{\mu}_0)'] \\ &= \mathbf{C}_0\boldsymbol{\Psi}_0\mathbf{C}_0' + \boldsymbol{\Theta}_0 , && (5.30) \end{aligned}$$

where $\boldsymbol{\Psi}_0 \equiv E(\mathbf{w}\ \mathbf{w}')$ and $\boldsymbol{\Theta}_0 \equiv E(\mathbf{v}\ \mathbf{v}')$.

The loglikelihood function then becomes

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{Y}) = & -\frac{N}{2}\log \mid \boldsymbol{\Sigma}_0 \mid -\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_{0i} - \boldsymbol{\mu}_{0i})'\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}_{0i} - \boldsymbol{\mu}_{0i}) \\ & -\frac{p_0 N}{2}\log 2\pi , \end{aligned} \tag{5.31}$$

where both $\mathbf{y}_{0i}$ and $\boldsymbol{\mu}_{0i}$ have subscript i because $\mathbf{u}$ may vary over subjects. In contrast to $\boldsymbol{\mu}_{0i}$, however, $\boldsymbol{\Sigma}_0$ is assumed to be common to all subjects. In fact, each subject in the sample is considered to be drawn from one of $N' \leq N$ distinct but, apart from the specific $\mathbf{u}$, equally distributed populations, having in particular $E(\mathbf{x}_{t_0})$, $\boldsymbol{\Phi}_{t_0}$ and all other parameter values equal. If q is the number of (fixed) elements in $\mathbf{u}$, the set of N' fixed values $\mathbf{u}_i$ in the sample must contain at least q linearly independent ones. An important advantage of $\mathbf{u}$ being fixed is that no distribution needs to be specified for its elements, which even need not be interval scale variables (e.g. income) but may also be dummy variables representing just group membership (e.g. gender).

While SEM programs do not maximize the loglikelihood function as given in Equation 5.31, it can, however, be proven (see Oud, in press, appendix A) that minimizing the SEM maximum likelihood fit function

$$F_{ML} = \log \mid \boldsymbol{\Sigma} \mid + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log \mid \mathbf{S} \mid - p , \tag{5.32}$$

for $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$

$$= \begin{bmatrix} \boldsymbol{\Phi}_u & \boldsymbol{\Phi}_u\mathbf{D}_0' \\ \mathbf{D}_0\boldsymbol{\Phi}_u & \mathbf{D}_0\boldsymbol{\Phi}_u\mathbf{D}_0' + \boldsymbol{\Sigma}_0 \end{bmatrix} ,$$

on the basis of the sample augmented moment matrix,

$$\mathbf{S}_{(p_0+q)\times(p_0+q)} = \frac{1}{N}\mathbf{Y}\ \mathbf{Y}' = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i\ \mathbf{u}_i' & \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i\ \mathbf{y}_{0i}' \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_{0i}\ \mathbf{u}_i' & \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_{0i}\ \mathbf{y}_{0i}' \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_u & \mathbf{S}_{u,y_0} \\ \mathbf{S}_{y_0,u} & \mathbf{S}_{y_0} \end{bmatrix},$$

gives the same result as maximizing Equation 5.31. In fact, $-\frac{N}{2}$ times the SEM fit function (Equation 5.32) is equal to the loglikelihood function (Equation 5.31) plus a constant.

5.5 The Kalman filter and its relationship to two cross-sectional estimators

The optimal or minimum variance estimator of the latent state $\mathbf{x}_t$, given past and current observations in $\mathbf{y}_0$, leads to the conditional mean or Kalman filter (Lewis, 1986, p. 11; Otter, 1984, p. 61)

$$\hat{\mathbf{x}}_t = E(\mathbf{x}_t|[\mathbf{y}_{t_0}'\mathbf{y}_{t_0+1}' \cdots \mathbf{y}_t']') ,$$

with error covariance matrix

$$\mathbf{P}_t = E[(\mathbf{x} - \hat{\mathbf{x}}_t)(\mathbf{x} - \hat{\mathbf{x}}_t)'|[\mathbf{y}_{t_0}'\mathbf{y}_{t_0+1}' \cdots \mathbf{y}_t']'] ,$$

The filter equations consist of a measurement update

$$\begin{aligned} \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t-} + \mathbf{H}_t(\mathbf{y}_t - \hat{\mathbf{y}}_{t-}) \quad \text{with} \quad \hat{\mathbf{y}}_{t-} = \mathbf{C}_t\hat{\mathbf{x}}_{t-} + \mathbf{D}_t\mathbf{u}_t , &(5.33)\\ \mathbf{P}_t &= (\mathbf{P}_{t-}^{-1} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1} , &(5.34) \end{aligned}$$

which represents the effects of the measurements $\mathbf{y}_t$, and a time update

$$\begin{aligned} \hat{\mathbf{x}}_{t-} &= \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} , &(5.35)\\ \mathbf{P}_{t-} &= \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}' + \mathbf{Q}_{t-1} , &(5.36) \end{aligned}$$

which represents the effects of the system dynamics. The filter, which is recursively based on past observations $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{t_0}$, has a predictor-corrector structure. The correction depends on the 'innovation' $\mathbf{y}_t - \hat{\mathbf{y}}_{t-} = \mathbf{y}_t - \mathbf{C}_t\hat{\mathbf{x}}_{t-} - \mathbf{D}_t\mathbf{u}_t = \mathbf{C}_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t-}) + \mathbf{v}_t$ which is weighted by the Kalman gain matrix $\mathbf{H}_t$ defined as

$$\mathbf{H}_t = \mathbf{P}_t\mathbf{C}_t'\mathbf{R}_t^{-1} . \qquad (5.37)$$

While the recursive computation of the filter estimates $\hat{\mathbf{x}}_t$ requires knowledge of the observations $\mathbf{y}_t$, this is not required for the recursive computation of the covariance matrices $\mathbf{P}_t$. The quality of the filtering results is known in advance before the actual processing of the observations takes place.

There exist many alternative formulations for $\mathbf{H}_t$ and $\mathbf{P}_t$. When $\mathbf{P}_{t-}$ is singular (which is the case, for example, when the successive $\mathbf{Q}_t$ matrices have diagonal zeroes as in the model of Equations 5.4 and 5.5) or $\mathbf{R}_t$ is singular (for example, when some of the observed variables are without measurement error), the following formulations are useful:

$$\begin{aligned} \mathbf{H}_t &= \mathbf{P}_{t-}\mathbf{C}_t'(\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t)^{-1} , & (5.38) \\ \mathbf{P}_t &= (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)\mathbf{P}_{t-}(\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)' + \mathbf{H}_t\mathbf{R}_t\mathbf{H}_t' . & (5.39) \end{aligned}$$

Equation 5.38 shows that only one matrix, the innovations covariance matrix $\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t$, must be inverted. This matrix may be invertible even when $\mathbf{P}_{t-}$ or $\mathbf{R}_t$ are singular.

Consideration of cross-sectional estimators is important in view of the initialization of the Kalman filter: how to start the Kalman filter recursion at the initial time point t_0 when no past observations are available? Because of the small number of time points typically used in behavioral science, the initial values $\hat{\mathbf{x}}_{t_0}$ and $\mathbf{P}_{t_0}$ do matter and must be chosen carefully. On the basis of their properties one could choose a cross-sectional estimator as initial estimator instead of starting the recursion with arbitrary values as is often done in control engineering. Depending on how the absent past is mathematically defined, that is, the way past time points $t' < t$ do not convey information about the state at time t, the Kalman filter reduces to the regression or to the Bartlett estimator (Lawley & Maxwell, 1971, p. 109-110).

If it is specified $\mathbf{A}_{t-1} = \mathbf{0}$ and, as in Equations 5.21-5.22, $E(\mathbf{x}_t) = \mathbf{B}_{t-1}\mathbf{u}_{t-1}$, one derives $\hat{\mathbf{x}}_{t-} = E(\mathbf{x}_t)$ and $\mathbf{P}_{t-} = E([\mathbf{x}_t - E(\mathbf{x}_t)][\mathbf{x}_t - E(\mathbf{x}_t)]') = \mathbf{\Phi}_t$. Then the Kalman filter reduces to the regression estimator (Jansen & Oud, 1995), and its error covariance matrix to the regression error covariance matrix (cfr. Equations 5.33 and 5.34):

$$\begin{aligned} \hat{\mathbf{x}}_t &= E(\mathbf{x}_t) + \mathbf{H}_t[\mathbf{y}_t - \mathbf{C}_tE(\mathbf{x}_t) - \mathbf{D}_t\mathbf{u}_t] \\ &= (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)E(\mathbf{x}_t) + \mathbf{H}_t(\mathbf{y}_t - \mathbf{D}_t\mathbf{u}_t) , & (5.40) \\ \mathbf{P}_t &= (\mathbf{\Phi}_t^{-1} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1} \\ &= \mathbf{\Phi}_t(\mathbf{I} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t\mathbf{\Phi}_t)^{-1} . & (5.41) \end{aligned}$$

For the case of $E(\mathbf{x}_t) = \mathbf{0}$ for all t, Equations 5.40 and 5.41 reduce to those of the regression estimator known from factor analysis (see Lawley & Maxwell, 1971, p. 109),

$$\begin{aligned} \hat{\mathbf{x}}_{tR} &= \mathbf{H}_t\mathbf{y}_t , & (5.42) \\ \mathbf{P}_{tR} &= \mathbf{\Phi}_t(\mathbf{I} + \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t\mathbf{\Phi}_t)^{-1} , & (5.43) \end{aligned}$$

where $\mathbf{P}_{tR}$ equals $\mathbf{P}_t$ in Equation 5.41.

However, in assuming $\mathbf{Q}_{t-1} \to \infty$, $\mathbf{P}_{t-} \to \infty$ (Equation 5.36), and $\mathbf{P}_t \to (\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1}$ (Equation 5.34). Then the weighting matrix $\mathbf{H}_t$ (Equation 5.37) becomes $(\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1}\mathbf{C}_t'\mathbf{R}_t^{-1}$, causing the Kalman filter equations (Equations 5.33 and 5.34) to reduce to those of the Bartlett estimator (see Lawley & Maxwell, 1971, p. 109):

$$
\begin{aligned}
\hat{\mathbf{x}}_{t_B} &= \hat{\mathbf{x}}_{t-} + (\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1}\mathbf{C}_t'\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{C}_t\hat{\mathbf{x}}_{t-}) \\
&= (\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1}\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{y}_t \ , \qquad (5.44) \\
\mathbf{P}_{tB} &= (\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t)^{-1} \ . \qquad (5.45)
\end{aligned}
$$

The interpretation of $\mathbf{Q}_{t-1} \to \infty$ is that the information from the past is completely unreliable, so that the filter utilizes present information only. Suppose on the other hand, that $\mathbf{R}_t^{-1} \to \mathbf{0}$, then $\mathbf{P}_t \to \mathbf{P}_{t-}$ because of $\mathbf{P}_{tB}^{-1} \to \mathbf{0}$ (cfr. Equation 5.34). Now $\mathbf{H}_t \to \mathbf{0}$ (Equations 5.37 and 5.38) and the Kalman estimator (Equation 5.33) utilizes past information only because of the present information being completely unreliable (Lewis, 1986, p. 77).

Calling inverse $\mathbf{P}_t^{-1}$ the information matrix, it becomes clear from Equation 5.34, that the information given by the Kalman filter is simply the sum of the information from two other estimators: the predictor $\hat{\mathbf{x}}_{t-}$ with information matrix $\mathbf{P}_{t-}^{-1}$ and the cross-sectional Bartlett estimator with information matrix $\mathbf{P}_{tB}^{-1} = \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t$.

The Kalman filter is derived as optimal in the sense that the error covariance matrix of any other linear estimator exceeds the error covariance matrix of the Kalman filter estimator by a non-negative definite matrix (Otter, 1984, p. 61). Because $\mathbf{P}_{t-}^{-1}$ in Equation 5.34 exceeds $\mathbf{\Phi}_t^{-1}$ in Equation 5.41, (assuming $\mathbf{P}_{t_0}^{-1}$ exceeds $\mathbf{\Phi}_{t_0}^{-1}$), and $\mathbf{\Phi}_t^{-1}$ exceeds $\mathbf{0}$ (Equation 5.45) with non-negative definite matrices, it is clear not only that the Kalman filter has less variance than each of the cross-sectional estimators, but also that the regression estimator has less variance than the Bartlett estimator. From this point of view the regression estimator is preferable to the Bartlett estimator for initializing the Kalman filter. Choosing the regression estimator as initial estimator and thus $\mathbf{P}_{t_0} = \mathbf{P}_{t_{0R}}$ also guarantees $\mathbf{P}_{t_0}^{-1}$ to exceed $\mathbf{\Phi}_{t_0}^{-1}$ and thus the Kalman filter to have least variance of the three estimators.

However, there are more criteria on the basis of which to choose an estimator than estimation error variance alone. As a second important criterium unbiasedness is considered. The weakest form is unconditional or populationwise unbiasedness:

$$E(\hat{\mathbf{x}}_t) = E(\mathbf{x}_t) \ . \qquad (5.46)$$

The strongest form is conditional or individualwise unbiasedness:

$$E(\hat{\mathbf{x}}_t|\mathbf{x}_t) = \mathbf{x}_t \ , \qquad (5.47)$$

(see Lawley & Maxwell, 1971, p. 108), making sure that the estimator $\hat{\mathbf{x}}_t$ does not only hit the true mean $E(\mathbf{x}_t)$ over the population of individuals considered, but also (over repeated measurements) the true value $\mathbf{x}_t$ for each individual or group of individuals with the same true value $\mathbf{x}_t$. This is important, because the estimator, although unconditionally unbiased, could perform very poorly (conditionally) for extreme high and extreme low true values. An intermediate form is t_0-conditional or modelwise unbiasedness:

$$E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0}) \ . \tag{5.48}$$

Here, for individuals with true value $\mathbf{x}_{t_0}$ at the initial time point t_0, the estimator is required to hit the predicted value $E(\mathbf{x}_t|\mathbf{x}_{t_0})$, to which the value $\mathbf{x}_{t_0}$ regresses (or egresses from), according to the true underlying dynamic model. According to the basic linear SSM (see Equation 5.1), for example, the predicted value is $E(\mathbf{x}_t|\mathbf{x}_{t_0}) = \mathbf{A}_{t-1}\mathbf{A}_{t-2}\ldots\mathbf{A}_{t_0}\mathbf{x}_{t_0}$.

As is proven below, the regression estimator turns out to be unbiased in none of the three senses, while the Bartlett estimator is unbiased in all of the three senses. The Kalman filter turns out to be t_0-conditionally as well as unconditionally unbiased, provided the initial estimator is conditionally unbiased. For proving these statements, the cross-sectional part of the Kalman filter (Equation 5.33) is rewritten as follows, assuming for convenience $\mathbf{D}_t\mathbf{u}_t = \mathbf{0}$ or writing $\mathbf{y}_t - \mathbf{D}_t\mathbf{u}_t$ as $\mathbf{y}_t$.

$$\begin{aligned}
\hat{\mathbf{x}}_t &= \mathbf{H}_t\mathbf{y}_t = \mathbf{H}_t(\mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t) = \mathbf{P}_t\mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t\mathbf{x}_t + \mathbf{H}_t\mathbf{v}_t \\
&= \mathbf{P}_t\mathbf{P}_{tB}^{-1}\mathbf{x}_t + \mathbf{H}_t\mathbf{v}_t \ .
\end{aligned} \tag{5.49}$$

Taking the expectations of Equations 5.46, 5.47, and 5.48:

$$\begin{aligned}
E(\hat{\mathbf{x}}_t) &= \mathbf{P}_t\mathbf{P}_{tB}^{-1}E(\mathbf{x}_t) \ , \\
E(\hat{\mathbf{x}}_t|\mathbf{x}_t) &= \mathbf{P}_t\mathbf{P}_{tB}^{-1}\mathbf{x}_t \ , \\
E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) &= \mathbf{P}_t\mathbf{P}_{tB}^{-1}E(\mathbf{x}_t|\mathbf{x}_{t_0}) \ ,
\end{aligned}$$

it follows immediately that the Bartlett estimator ($\mathbf{P}_t = \mathbf{P}_{tB}$) is intrinsically unbiased in all three senses and the regression estimator ($\mathbf{P}_t = \mathbf{P}_{tR} \neq \mathbf{P}_{tB}$) in none of them. Of course, the regression estimator is unconditionally unbiased in assuming that $E(\mathbf{x}_t) = \mathbf{0}$, or by compensating the unconditional bias through the addition of a corresponding correction term $(\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)E(\mathbf{x}_t) = (\mathbf{I} - \mathbf{P}_t\mathbf{P}_{tB}^{-1})E(\mathbf{x}_t)$ (cfr. Equation 5.40 which is a more general version of Kelley's true score estimator, Lord & Novick, 1968, p. 152). This, however, does not take away the regression estimator's conditional as well as t_0-conditional bias towards the mean trajectory, causing serious underestimation for high-achievers and serious overestimation for low-achievers.

Because the Bartlett estimator has minimum variance in the class of unbiased estimators (Lawley & Maxwell, 1971 pp 110-111), it seems preferable to the regression estimator for initializing the Kalman filter Note that although the regression estimator has less variance than the Bartlett estimator and thus at t_0 $\mathbf{P}_{t_{0B}}$ exceeds $\mathbf{P}_{t_{0R}}$, immediately at the next time point t_0+1 the Bartlett initialized Kalman filter already performs better than the regression estimator (assuming $\mathbf{\Phi}_{t_0}$ exceeds $\mathbf{P}_{t_{0B}}$ but this will be the case except for extremely high measurement error variances in $\mathbf{R}_{t_0}$) It should be noted also, that while the Kalman filter is based on the orthogonality principle, meaning that $E(\mathbf{e}_t\ \hat{\mathbf{x}}_t') = \mathbf{0}$, the Bartlett estimator has $E(\mathbf{e}_t\ \mathbf{x}_t') = \mathbf{0}$ but $E(\mathbf{e}_t\ \hat{\mathbf{x}}_t') \neq \mathbf{0}$ Under rather mild conditions, however, it can be proven that $E(\mathbf{e}_t\ \hat{\mathbf{x}}_t') \to \mathbf{0}$ as time proceeds, implying that eventually the Bartlett initialized filter again gets the same covariance matrix as the regression initialized filter (see Oud et al , 1990, pp 407 408) Also it can be proven, that when initializing by means of the Bartlett estimator the Kalman filter recursion in Equations 5 38-5 39 remains valid

The proof of the t_0-conditional unbiasedness of the Kalman filter is recursive with two substeps in each recursion 1 if the previous filter $\hat{\mathbf{x}}_{t-1}$ is t_0-conditionally unbiased, then the predictor $\hat{\mathbf{x}}_{t-}$, 2 if the predictor $\hat{\mathbf{x}}_{t-}$ is t_0 conditionally unbiased, then the new filter $\hat{\mathbf{x}}_t$ Thus, for the Kalman filter to be t_0-conditionally unbiased as a whole, the recursion must start with an initial estimator $\mathbf{x}_{t_0}$ which is t_0 conditionally unbiased, that is $E(\hat{\mathbf{x}}_{t_0}|\mathbf{x}_{t_0}) = E(\mathbf{x}_{t_0}|\mathbf{x}_{t_0}) = \mathbf{x}_{t_0}$, which means that the initial estimator must be conditionally unbiased The proof of substep 1 uses the equalities

$$\begin{aligned} E(\hat{\mathbf{x}}_{t-}|\mathbf{x}_{t_0}) &= E(\mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1}|\mathbf{x}_{t_0}) , \\ E(\mathbf{x}_t|\mathbf{x}_{t_0}) &= E(\mathbf{A}_{t-1}\mathbf{x}_{t-1}|\mathbf{x}_{t_0}) \end{aligned}$$

Introducing the premise $E(\hat{\mathbf{x}}_{t-1}|\mathbf{x}_{t_0}) = E(\mathbf{x}_{t-1}|\mathbf{x}_{t_0})$, the conclusion $E(\hat{\mathbf{x}}_{t-}|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0})$ follows immediately For substep 2 we take the t_0-conditional expectation of both sides of Equation 5 33

$$E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)E(\hat{\mathbf{x}}_{t-}|\mathbf{x}_{t_0}) + \mathbf{H}_tE(\mathbf{y}_t|\mathbf{x}_{t_0})$$

Introducing the premise $E(\hat{\mathbf{x}}_{t-}|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0})$ and using $\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t$ and $E(\mathbf{v}_t|\mathbf{x}_{t_0}) = \mathbf{0}$, one derives

$$E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)E(\mathbf{x}_t|\mathbf{x}_{t_0}) + \mathbf{H}_t\mathbf{C}_tE(\mathbf{x}_t|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0}) ,$$

which is the desired conclusion

5.5.1 Initialization of the filter on the basis of the state-trait model

Because in the state-trait model trait variables are added to the state vector (cfr. Equations 5.10 and 5.11), a specific problem turns up as regards the initialization of the filter. In considering the Bartlett estimator (Equations 5.44 and 5.45), it is seen from Equation 5.11 that the factorloading matrix $\mathbf{C}_t = [\mathbf{C}_t^\circ\ \mathbf{0}]$ contains a zero columnvector. It leads to the Bartlett error covariance matrix $\mathbf{P}_{tB}$ becoming noninvertible, implying that it cannot be used for the initialization of the filter. The cross-sectional regression estimator (Equations 5.42 and 5.43), however, can be applied without any problems but, as has been shown, it is conditionally as well as t_0-conditionally biased. Therefore another approach is proposed, guaranteeing t_0-conditional unbiasedness.

Using Equation 5.38 for $\mathbf{C}_t = [\mathbf{C}_t^\circ\ \mathbf{0}]$, and

$$\mathbf{P}_{t-} = \begin{bmatrix} \mathbf{P}_{x_{t-}^\circ} & \mathbf{P}_{x_{t-}^\circ,\xi} \\ \mathbf{P}_{\xi,x_{t-}^\circ} & \mathbf{P}_\xi \end{bmatrix}, \tag{5.50}$$

(cfr. Equation 5.12) one derives

$$\mathbf{H}_t\mathbf{C}_t = \begin{bmatrix} \mathbf{P}_{x_t^\circ}\mathbf{C}_t^{\circ\prime}\mathbf{R}_t^{-1}\mathbf{C}_t^\circ & \mathbf{0} \\ \mathbf{P}_{\xi,x_{t-}^\circ}\mathbf{P}_{x_{t-}^\circ}^{-1}\mathbf{P}_{x_t^\circ}\mathbf{C}_t^{\circ\prime}\mathbf{R}_t^{-1}\mathbf{C}_t^\circ & \mathbf{0} \end{bmatrix}. \tag{5.51}$$

Then applying $\hat{\mathbf{x}}_t = \mathbf{H}_t\mathbf{y}_t$ as in Equation 5.49, but using the Bartlett estimator for the state vector $\mathbf{x}_t^\circ$ only and thus substituting $\mathbf{P}_{tB}^\circ = (\mathbf{C}_t^{\circ\prime}\mathbf{R}_t^{-1}\mathbf{C}_t^\circ)^{-1}$ for $\mathbf{P}_{x_t^\circ}$ in Equation 5.51, it follows that

$$\mathbf{H}_t\mathbf{C}_t = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{\xi,x_{t-}^\circ}\mathbf{P}_{x_{t-}^\circ}^{-1} & \mathbf{0} \end{bmatrix}, \tag{5.52}$$

causing $\hat{\mathbf{x}}_t = \mathbf{H}_t\mathbf{y}_t = \mathbf{H}_t\mathbf{C}_t\mathbf{x}_t + \mathbf{H}_t\mathbf{v}_t$ in Equation 5.49 to become

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{x}}_t^\circ \\ \hat{\boldsymbol{\xi}} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{\xi,x_{t-}^\circ}\mathbf{P}_{x_{t-}^\circ}^{-1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^\circ \\ \boldsymbol{\xi} \end{bmatrix} + \mathbf{H}_t\mathbf{v}_t \\ &= \begin{bmatrix} \mathbf{x}_t^\circ \\ \mathbf{P}_{\xi,x_{t-}^\circ}\mathbf{P}_{x_{t-}^\circ}^{-1}\mathbf{x}_t^\circ \end{bmatrix} + \mathbf{H}_t\mathbf{v}_t\,. \end{aligned} \tag{5.53}$$

Then, at $t = t_0$ inserting $\mathbf{P}_{\xi,x_{t-}^\circ} = \boldsymbol{\Phi}_{\xi,x_{t_0}^\circ}$ and $\mathbf{P}_{x_{t-}^\circ} = \boldsymbol{\Phi}_{x_{t_0}^\circ}$ (see Equation 5.12) and taking the conditional expectation (see Equation 5.47), one derives for $\hat{\mathbf{x}}_{t_0}^\circ$

$$E(\hat{\mathbf{x}}_{t_0}^\circ | \mathbf{x}_{t_0}^\circ, \boldsymbol{\xi}) = \mathbf{x}_{t_0}^\circ\ ,$$

and for

$$
\begin{aligned}
\hat{\boldsymbol{\xi}} &= \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \hat{\mathbf{x}}^\circ_{t_0} \\
&= \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{H}^\circ_{t_0} \mathbf{y}_{t_0} \\
&= \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{H}^\circ_{t_0} \mathbf{C}^\circ_{t_0} \mathbf{x}^\circ_{t_0} + \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{H}^\circ_{t_0} \mathbf{v}_{t_0} \\
&= \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{x}^\circ_{t_0} + \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{H}^\circ_{t_0} \mathbf{v}_{t_0} \ ,
\end{aligned}
$$

$$
E(\hat{\boldsymbol{\xi}}|\mathbf{x}^\circ_{t_0}, \boldsymbol{\xi}) = \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \mathbf{x}^\circ_{t_0} \ ,
$$

which for multinormally distributed variables turns out to be equal to the linear regression problem of $\boldsymbol{\xi}$ on $\mathbf{x}^\circ_{t_0}$:

$$
E(\hat{\boldsymbol{\xi}}|\mathbf{x}^\circ_{t_0}, \boldsymbol{\xi}) = E(\boldsymbol{\xi}|\mathbf{x}^\circ_{t_0}) \ .
$$

So, using the Bartlett estimator for estimating the state $\mathbf{x}^\circ_{t_0}$ at t_0 leads to a conditionally unbiased estimator of $\mathbf{x}^\circ_{t_0}$ as well as of $\boldsymbol{\xi}$ at t_0 (both conditional on $\mathbf{x}_{t_0}$) and next to t_0-conditional unbiasedness of the Kalman filter at later points in time, as an estimator of both the state and the trait. The accompanying initial error covariance matrix turns out to be

$$
\mathbf{P}_{t_0} = \begin{bmatrix} (\mathbf{C}^{\circ\prime}_{t_0} \mathbf{R}^{-1}_{t_0} \mathbf{C}^\circ_{t_0})^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_\xi - \boldsymbol{\Phi}_{\xi,x^\circ_{t_0}} \boldsymbol{\Phi}^{-1}_{x^\circ_{t_0}} \boldsymbol{\Phi}'_{\xi,x^\circ_{t_0}} \end{bmatrix} . \tag{5.54}
$$

5.6 The Kalman smoother and the 'overall' regression estimator

By making use of the normal-correlation theorem (Liptser & Shiryayev, 1978, chap. 13; see also Lewis, 1986, ex. 1.1-2), writing for convenience $t_0 + T - 1 = s$ and $\mathbf{y}_0 = \mathbf{y}$ (see page 77), the optimal estimator of the latent states in vector $\mathbf{x} = [\mathbf{x}'_{t_0} \mathbf{x}'_{t_0+1} \ldots \mathbf{x}'_s]'$ for observations $\mathbf{y} = [\mathbf{y}'_{t_0} \mathbf{y}'_{t_0+1} \ldots \mathbf{y}'_s]'$ is the conditional mean

$$
\begin{aligned}
\hat{\mathbf{x}} &= E(\mathbf{x}|\mathbf{y}) \\
&= E(\mathbf{x}) + \boldsymbol{\Phi}_{\mathrm{xy}} \boldsymbol{\Phi}^{-1}_{\mathrm{y}} [\mathbf{y} - E(\mathbf{y})] \ ,
\end{aligned} \tag{5.55}
$$

with conditional error covariance matrix

$$
\begin{aligned}
\mathbf{P} &= E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})'|\mathbf{y}] \\
&= \boldsymbol{\Phi} - \boldsymbol{\Phi}_{\mathrm{xy}} \boldsymbol{\Phi}^{-1}_{\mathrm{y}} \boldsymbol{\Phi}'_{\mathrm{xy}} \ ,
\end{aligned} \tag{5.56}
$$

for $\mathbf{\Phi}_{xy} = cov(\mathbf{xy})$, $\mathbf{\Phi}_y = cov(\mathbf{y})$, and $\mathbf{\Phi} = cov(\mathbf{x})$. This 'overall' regression estimator utilizes past, current and future information. The estimate of $\mathbf{x}_t$ in $\hat{\mathbf{x}}$ for $t \leq s$ is called the Kalman smoother (Meditch, 1969, p. 206) and is written as

$$\hat{\mathbf{x}}_t^s = E(\mathbf{x}_t | [\mathbf{y}'_{t_0} \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_s]') , \tag{5.57}$$

$$\mathbf{P}_t^s = E(\mathbf{e}_t^s \, \mathbf{e}_t^{s'} | [\mathbf{y}'_{t_0} \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_s]') , \tag{5.58}$$

for estimation error $\mathbf{e}_t^s = \mathbf{x}_t - \hat{\mathbf{x}}_t^s$. A special case of $\hat{\mathbf{x}}_t^s$ is $\hat{\mathbf{x}}_s^s$ or the (cross-sectional regression estimator initialized) Kalman filter.

By writing out Equations 5.55 and 5.56 on the basis of the overall output equation $\mathbf{y} = \overline{\mathbf{C}}\mathbf{x} + \overline{\mathbf{D}}\mathbf{u} + \mathbf{v}$; $\overline{\mathbf{C}}$ consisting of submatrices $\mathbf{C}_{t_0}$, $\mathbf{C}_{t_0+1}$, $\ldots$, $\mathbf{C}_s$; $\overline{\mathbf{D}}$ and $\mathbf{u}$, respectively, of $\mathbf{D}_{t_0}$, $\mathbf{D}_{t_0+1}$, $\ldots$, $\mathbf{D}_s$, and $\mathbf{u}_{t_0}$, $\mathbf{u}_{t_0+1}$, $\ldots$, $\mathbf{u}_s$, and $E(\mathbf{v}\ \mathbf{v}') = \overline{\mathbf{R}}$ consisting of submatrices $\mathbf{R}_{t_0}$, $\mathbf{R}_{t_0+1}$, $\ldots$, $\mathbf{R}_s$ on the diagonal (cfr. Equations 5.25-5.30), the estimator takes the following form (e.g. Lewis, 1986, ex. 1.1-3),

$$\hat{\mathbf{x}} = E(\mathbf{x}) + \mathbf{\Phi}\overline{\mathbf{C}}'(\overline{\mathbf{C}}\mathbf{\Phi}\overline{\mathbf{C}}' + \overline{\mathbf{R}})^{-1}[\mathbf{y} - \overline{\mathbf{C}}E(\mathbf{x})] , \tag{5.59}$$

$$\mathbf{P} = \mathbf{\Phi} - \mathbf{\Phi}\overline{\mathbf{C}}'(\overline{\mathbf{C}}\mathbf{\Phi}\overline{\mathbf{C}}' + \overline{\mathbf{R}})^{-1}\overline{\mathbf{C}}\mathbf{\Phi} \tag{5.60}$$

$$= (\mathbf{\Phi}^{-1} + \overline{\mathbf{C}}'\overline{\mathbf{R}}^{-1}\overline{\mathbf{C}})^{-1} . \tag{5.61}$$

The input-effects in $\overline{\mathbf{D}}\mathbf{u}$ are both in $\mathbf{y}$ and $E(\mathbf{y})$, meaning that $\mathbf{y} - \overline{\mathbf{C}}E(\mathbf{x})$ at the right-hand side of Equation 5.59, in fact, equals $\mathbf{y} - E(\mathbf{y})$. As Equations 5.59-5.61 involve complex operations, such as inverting very large matrices, it often is recommended (Molenaar & Oud, 1991; Singer, 1992) to use the numerically more efficient recursive Kalman smoother algorithm involving much smaller matrices than the 'overall' regression estimator. The off-diagonal matrices $\mathbf{P}_{t,t-k}^s$ for smoother errors $\mathbf{e}_t^s$ and $\mathbf{e}_{t-k}^s$ and $k \neq 0$, are given by Jansen and Oud (1995). Note that one gets the Kalman filter (Equations 5.33-5.34) from Equations 5.59-5.61 by taking $t = s$ in $\hat{\mathbf{x}}_t^s$ and the cross-sectional regression estimator by taking $t = s = t_0$.

The Kalman smoother can be seen as a backward filtering problem, starting at the point the forward filter finishes. A convenient scheme which incorporates the forward filter and the backward filter, is the Rauch-Tung-Striebel formulation (Rauch et al., 1965; also Lewis, 1986, chap. 2.8). It only uses the Kalman filter state estimates $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{x}}_{t-}$, and its error covariance matrices $\mathbf{P}_t$ and $\mathbf{P}_{t-}$. That is

$$\hat{\mathbf{x}}_t^s = \hat{\mathbf{x}}_t + \mathbf{F}_t(\hat{\mathbf{x}}_{t+1}^s - \hat{\mathbf{x}}_{(t+1)-}) , \tag{5.62}$$

$$\mathbf{P}_t^s = \mathbf{P}_t + \mathbf{F}_t(\mathbf{P}_{t+1}^s - \mathbf{P}_{(t+1)-})\mathbf{F}'_t , \tag{5.63}$$

with the smoother gain matrix being

$$\mathbf{F}_t = \mathbf{P}_t\mathbf{A}'_t(\mathbf{P}_{(t+1)-})^{-1} . \tag{5.64}$$

The Kalman smoother state estimate $\hat{\mathbf{x}}_t^s$ (Equation 5.62) and error covariance matrix $\mathbf{P}_t^s$ (Equation 5.63) require the Kalman filter estimate $\hat{\mathbf{x}}_t$ and covariance matrix $\mathbf{P}_t$ with $\hat{\mathbf{x}}_s$ and $\mathbf{P}_s$ as initial conditions (Jazwinski, 1970, p. 217; Lewis, 1986, p. 134; Rauch et al., 1965, p. 1447). Note that $\hat{\mathbf{x}}_s = \hat{\mathbf{x}}_s^s$ and $\mathbf{P}_s = \mathbf{P}_s^s$ (the Kalman filter coinciding with the Kalman smoother at $t = s$; no future information available), and that because of the Kalman smoother being equivalent to the 'overall' regression estimator, the Kalman filter estimates at $t = s$ coincide with the estimates of the 'overall' regression estimator at $t = s$ as well.

Also for the Kalman smoother it is crucial to consider t_0-conditional unbiasedness to make sure that inserting the Bartlett estimator at the starting point of the Kalman filter leads to t_0-conditional unbiasedness even in the backward smoother recursion. Because the smoother is initialized by the Kalman filter estimates, the initial smoother estimates can be assumed to be t_0 conditionally unbiased (see page 83). Then, for $t+1 = t_0+T-1 = s$, the smoother state equation (Equation 5.62) is written as

$$\hat{\mathbf{x}}_t^s = \hat{\mathbf{x}}_t + \mathbf{F}_t(\hat{\mathbf{x}}_{t+1} - \hat{\mathbf{x}}_{(t+1)-}) \; , \tag{5.65}$$

with initial condition $\hat{\mathbf{x}}_{t+1}^s = \hat{\mathbf{x}}_{t+1}$. Taking the conditional expectation,

$$E(\hat{\mathbf{x}}_t^s|\mathbf{x}_{t_0}) = E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) + \mathbf{F}_t[E(\hat{\mathbf{x}}_{t+1}|\mathbf{x}_{t_0}) - E(\hat{\mathbf{x}}_{(t+1)-}|\mathbf{x}_{t_0})] \; , \tag{5.66}$$

and decomposing the state estimates of Equation 5.65 as follows,

$$\begin{aligned}
\hat{\mathbf{x}}_t &= \mathbf{x}_t - \mathbf{e}_t \; , \\
\hat{\mathbf{x}}_{t+1} &= \mathbf{x}_{t+1} - \mathbf{e}_{t+1} \; , \\
\hat{\mathbf{x}}_{(t+1)-} &= \mathbf{x}_{t+1} - \mathbf{e}_{(t+1)-} \; ,
\end{aligned}$$

it follows that, because of $E(\mathbf{e}_t|\mathbf{x}_{t_0}) = E(\mathbf{e}_{t+1}|\mathbf{x}_{t_0}) = E(\mathbf{e}_{(t+1)-}|\mathbf{x}_{t_0}) = \mathbf{0}$,

$$E(\hat{\mathbf{x}}_t^s|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0}) + \mathbf{F}_t[E(\mathbf{x}_{t+1}|\mathbf{x}_{t_0}) - E(\mathbf{x}_{t+1}|\mathbf{x}_{t_0})] \; , \tag{5.67}$$

or

$$E(\hat{\mathbf{x}}_t^s|\mathbf{x}_{t_0}) = E(\hat{\mathbf{x}}_t|\mathbf{x}_{t_0}) = E(\mathbf{x}_t|\mathbf{x}_{t_0}) \; , \tag{5.68}$$

which proves the t_0-conditionally unbiasedness of $\hat{\mathbf{x}}_t^s$ (cfr. Equation 5.48). As $\hat{\mathbf{x}}_t^s$ is t_0-conditionally unbiased, recursively the smoother state estimates $\hat{\mathbf{x}}_{t-1}^s, \hat{\mathbf{x}}_{t-2}^s, \ldots, \hat{\mathbf{x}}_{t_0}^s$ in the backward recursion can be proven to be t_0-conditionally unbiased also. Moreover, at t_0, $\hat{\mathbf{x}}_{t_0}^s$ turns out to be as conditionally unbiased as the Bartlett estimator, because $E(\hat{\mathbf{x}}_{t_0}^s|\mathbf{x}_{t_0}) = \mathbf{x}_{t_0}$, but with less variance than the Bartlett estimator.

5.7 Two examples

Both the Kalman filter and Kalman smoother can be applied on the basis of the SEM state space model. First, it is shown how filtering and smoothing is to be performed on the basis of the structured means SEM model, assuming the input to consist of the unit input-variable only and no trait variables to be present in the model. An adapted version of the zero means Beginning Reading model (Oud et al., 1990) is discussed with respect to model construction and parameter identification. Second, filtering and smoothing is performed on the basis of the state-trait model. The adapted version of the Kalman filter estimator is used with conditionally unbiased initial estimates of both the states and traits and yielding t_0-conditional unbiased estimates of the states and traits at later time points.

5.7.1 The structured means SEM model: an example

The structured means SEM model assumes that the input-vector $\mathbf{u}$ of the SEM model in Equations 5.25 and 5.26 consists of the unit input ($u_t = 1$ for all t) only. The associated subvectors $\mathbf{b}_{t-1}$ in $\overline{\mathbf{b}}$ and $\mathbf{d}_t$ in $\overline{\mathbf{d}}$ contain the intercepts which make it possible to model absolute latent growth on the basis of observed growth. The $\mathbf{d}_t$ coefficients are central in keeping information about the origins of the measurement instrument scales in relation to the latent scales in the model. The $\mathbf{b}_{t-1}$ coefficients define mean latent growth. If $\overline{\mathbf{b}} = \mathbf{0}$, $E(\mathbf{x}) = \mathbf{0}$, and the means, $E(\mathbf{y}_0) = \overline{\mathbf{d}}$, are typically left unconstrained in the model. Restricting the $\mathbf{d}_t$ coefficients appropriately, however, and allowing $\mathbf{b}_{t-1} \neq \mathbf{0}$, nonzero mean development, $E(\mathbf{x}) \neq \mathbf{0}$, is defined.

The estimation of absolute latent developmental curves relates to the identification procedures of the longitudinal SEM model, requiring the scales of the latent variables to maintain the same unit and origin over the entire time range. This must be the case if absolute latent growth is to make sense at all. As the $\mathbf{C}_t$ coefficients are central in determining the latent standard deviations and thus the latent scale units and the $\mathbf{d}_t$ coefficients in combination with the $\mathbf{b}_{t-1}$ coefficients in determining the latent means and thus the latent scale origins, it is clear that for identification arbitrary restrictions on the $\mathbf{C}_t$ and $\mathbf{d}_t$ coefficients are not allowed, except at the initial point in time. For identification in an absolute growth model, all $\mathbf{C}_t$ and $\mathbf{d}_t$ coefficients at later points in time must be linked in some way to those at the initial point in time.

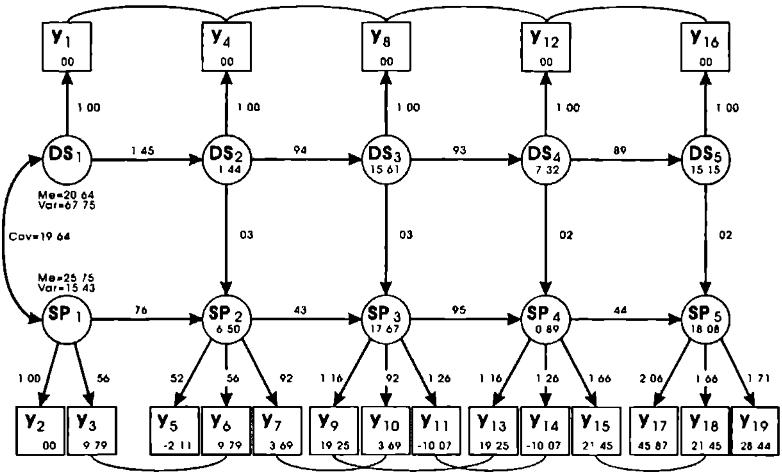
First, for measuring a particular latent variable over time, one or several identical instruments can be used at each time point. The equality restrictions over time between the $\mathbf{d}_t$ coefficients and $\mathbf{C}_t$ coefficients of a single instrument generally suffice for the identification of the $\mathbf{b}_{t-1}$ coefficients as well as the latent standard deviations of a single underlying latent variable. This situation, however, often is no solution to the identification problem. Many past longitudinal research projects in behavioral science failed, just because the measuring instruments chosen at the start turned out to have insufficient ceiling at later developmental stages and became useless. In many cases the identification problem can be solved more realistically in the following way. Only at pairs of successive points in time the same measuring instruments are chosen: instrument A applied at times 1 and 2, instrument B at times 2 and 3, instrument C at times 3 and 4, etc. Between pairs of successive time points the $\mathbf{C}_t$ and $\mathbf{d}_t$ coefficients of the same instrument must be specified equal over time. At the same time point the $\mathbf{C}_t$ and $\mathbf{d}_t$ coefficients of different instruments are linked to one another by the common latent variable.

That different instruments at the same time point indeed measure the same latent content can be checked by Jöreskog's congenericity test (Jöreskog, 1974, pp. 5-12). Congeneric instruments measure the same underlying variable, which is meant in the sense that the underlying latent components of the observed variables correlate 1. Although congenericity imposes rather strong requirements on the instruments involved, the measurement of the same underlying variables may nevertheless be in different $\mathbf{C}_t$ coefficients or observed scale units, different $\mathbf{d}_t$ coefficients or observed scale origins, also with different reliabilities, and a different number of observed variables. Congenericity is less restrictive than parallelism, τ-equivalence or essential τ-equivalence of instruments (Lord & Novick, 1968, pp. 47-50).

Both identification procedures were applied to the adapted version of the Beginning Reading model in Figure 5.1 (see Oud, et al., 1990). For a sample of $N = 225$ primary school pupils the latent decoding speed (DS) and spelling (SP) development was modeled over five measurement time points. Figure 5.1 shows all nonzero fixed and estimated free parameters values in $\boldsymbol{B}$ and $\boldsymbol{\Lambda}$ (cfr. Equations 5.25 and 5.26). The latent intercept values $\mathbf{b}_{t-1}$ are given in the circles of the latent variables (except for the initial latent means 20.64 and 25.75 of DS_1 and SP_1 in $\mathbf{b}_{t_0-1}$), while the values of all $\mathbf{d}_t$ coefficients are given in the squares of the observed variables. Parameter values in the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ are not given (except for the initial non-explained latent variances 67.75 and 15.43 and initial covariance 19.64 of DS_1 and SP_1 in $\boldsymbol{\Psi}$). These matrices were specified diagonal (except for the initial covariance between DS_1 and SP_1 in $\boldsymbol{\Psi}$).

The autoregressive effects for the DS-variables are much higher than for the SP-variables, except between the third and fourth time point in the model. Furthermore, autoregressive effects are much higher than the instantaneous effects between the DS- and SP-variables. As regards identification of the $\mathbf{d}_t$, $\mathbf{C}_t$, and $\mathbf{b}_{t-1}$ coefficients, the first situation applied to the DS part of the model. The $\mathbf{d}_t$ and $\mathbf{C}_t$ coefficients of the same observed variables y_1, y_4, y_8, y_{12}, and y_{16} were constrained to be equal over time (this is indicated by the curved lines in Figure 5.1). In additionally fixing $d_1 = .00$ and $c_{1,DS_1} = 1.00$ for y_1 at the initial time point, all parameters of DS in $\overline{\mathbf{b}}$ became identified. Also its initial mean in $\mathbf{b}_{t_0-1}$ became equal to the mean observed value of y_1 (20.64), and the latent variance of

Fig. 5.1: Structured means SEM model for Beginning Reading (nonstandardized solution).



DS_1 in $\boldsymbol{\Psi}$ became equal to the true variance of y_1. Other initial value restrictions could have been chosen and would have led to different values for the initial latent mean and variance (e.g., zero mean and unit variance), but it seems advantageous to be able to interpret the absolute latent development in terms of the mean and true standard deviation of a real instrument at the start.

In the same way, with regard to the SP part of the model, the latent mean and variance of SP_1 at the initial time point (25.75 and 15.43, respectively) were provided by the spelling instrument used in y_2 by means of restrictions $d_2 = .00$ and $c_{2,SP_1} = 1.00$. Apart from these initial value restrictions, however, the quite different identification procedure of situation two was followed. Different instruments were used over time, but with overlap between successive times as indicated by the five curved lines at the bottom of Figure 5.1. The corresponding pairs of equality constraints between the $\mathbf{d}_t$ and $\mathbf{C}_t$ coefficients were introduced and led to the identification of all $\mathbf{b}_{t-1}$ coefficients and indirectly determined the latent means and latent standard deviations as given.

The 5 pairs of equality constraints for the observed SP variables together with the fixations at the initial time point formed 6 pairs of restrictions, one more pair than necessary for identification: going from time point 3 to time point 4 two instruments were used twice (i.e. y_9 and y_{13}, and y_{11} and y_{14}, respectively, came from the same instrument). The two overidentifying restrictions could be tested by the usual χ^2-difference test of the SEM program: $\chi^2 = \chi^2_{141} - \chi^2_{139} = .4$, which for *df* = 2 is not significant. Also, on the basis of the estimated $\mathbf{\Lambda}$ and $\mathbf{\Theta}$, the correlations between the latent components of the observed variables were computed in a subsequent SEM analysis. The 13 congenericity correlations between observed SP variables at identical time points (between y_2 and y_3, between y_5, y_6 and y_7, etc.) turned out to be all at least .974 and, when restricted to the value of 1, not to differ significantly from that value: $\chi^2 = 4.02$ for *df* = 13. The tests and the uniformly high congenericity correlations confirmed that the latent SP variables keep the same content over time and that the absolute latent development may be interpreted in terms of this same content.

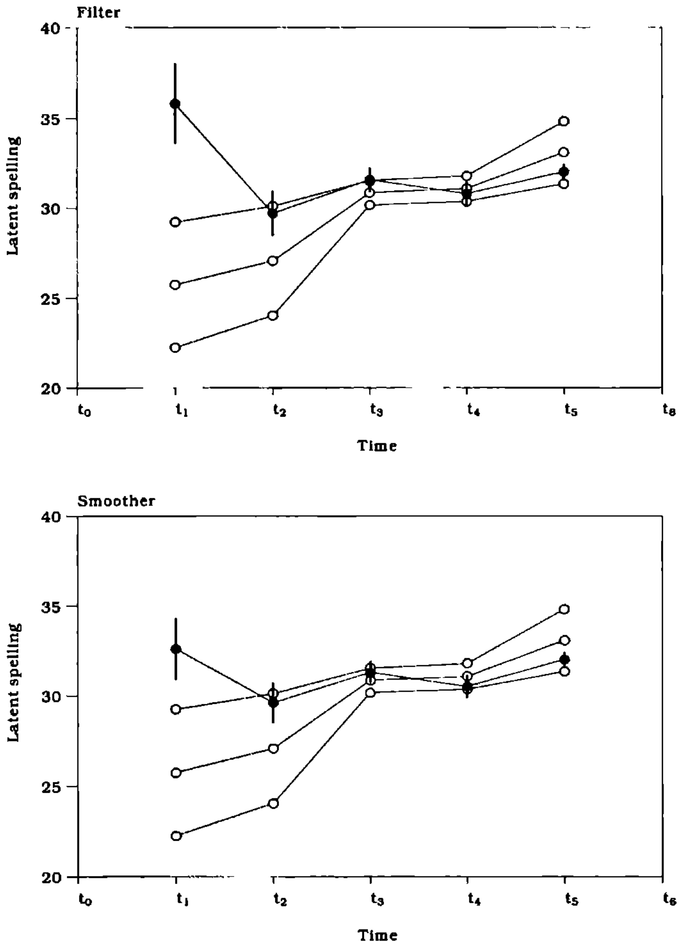
Figure 5.2 gives an example of the filter and smoother estimates. Because in addition to the pupil's estimates with associated standard errors (square roots of the diagonals of the Kalman covariance matrices), the population mean ± 1 standard deviation is shown at each of the 5 time points, the graphs allow the absolute information to be interpreted relatively, in relation to the population's development, as well. While absolutely the pupil's spelling ability is decreasing only between time points 1 and 2 and between time points 3 and 4, relatively the position of the pupil in the population goes down over the whole time range. The relatively large standard error at the initial time point results from the Bartlett estimator, which was used as initial estimator for the Kalman filter. The Kalman filter estimates at later time points, which are based on more and more information, show a substantial increase in precision. The interpretation of the smoother estimates is practically the same, except that the developmental curve is smooth (no sudden updates at observation time points) and that its associated standard errors for the first four points in time are somewhat smaller.

5.7.2 The state-trait SEM model: an example

Filtering and smoothing can also be performed on the basis of a structured means SEM model which includes trait variables (cfr. Equations 5.19-5.20). Within an educational context a trait or constant subject specific intercept characterizes a pupil's general ability level for the schoolperiod to which the state-trait model applies. It gives a constant contribution to the ability development and is to be compared to a zero mean over the population of pupils.

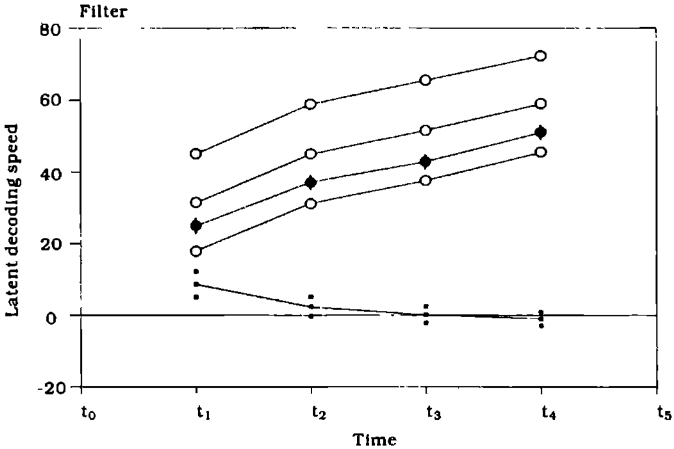
A state-trait model, including the unit input-variable, was estimated for decoding speed (DS) on the basis of a sample of $N = 740$ primary school pupils from grade 4 to grade 5. The model consisted of four measurement time points. Because of the constancy over time the DS trait variable only needed to be specified once in the SEM model. The regression effects of the states (DS_2, DS_3, DS_4) on the trait were fixed to unity whereas the variance of the trait and its covariance with the initial state DS_1 were specified to be free. The SEM output indicated that both the trait variance (35.05) and initial state-trait covariance (64.38) differed from zero and were significant ($\alpha = .05$).

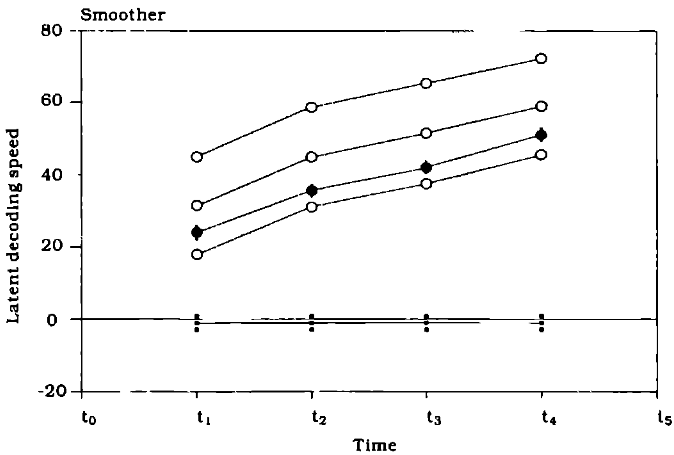
Fig. 5.2: A pupil's Kalman filter and Kalman smoother estimates (black circles) of absolute latent spelling scores with associated standard errors (vertical lines encompassing estimates ±1 standard error) in addition to the population's means ±1 standard deviation (white circles).



Subject specific intercept terms were estimated by application of the Kalman filter and Kalman smoother, and by making use of the initial estimate as proposed in section 5.5.1. Figure 5.3 shows the filter and smoother estimates of the latent decoding speed development and of the associated trait value of one pupil. The filter estimates of the decoding speed development show an absolute increase over time, but relatively the pupil's level keeps the same over the four time points. The

Fig. 5.3: A pupil's Kalman filter and Kalman smoother estimates (black circles) of absolute latent decoding speed scores with associated standard errors (vertical lines encompassing estimates ±1 standard error), and its associated trait value with ±1 standard error (black squares), in addition to the population's means ±1 standard deviation (white circles).





smoother decoding speed estimates, in fact, yield the same picture. At each time point, the Kalman filter processes additional information for the estimation of the constant trait value, leading to a more precise estimate. At the final time point, when all information has been processed, there is no precision gain left in the estimation of the trait value. It results in the subject specific estimate to be unaltered in the backward recursion, meaning that the Kalman smoother does not improve upon the results of the Kalman filter. The subject specific trait effect is just below the population zero mean value of the trait variable. It confirms that the pupil's average achievement level for decoding speed is below the population's absolute mean level.

5.8 Discussion

To finish three problems as yet unsolved in the proposed Kalman filter and smoother procedure are mentioned: (a) the way the sampling variability of the SEM parameter estimates can be systematically accounted for in the standard errors of the Kalman filter and smoother estimates, (b) how to derive SSM based standard errors for means of Kalman filter and smoother estimates over groups of individuals, (c) the construction of a statistical test to evaluate whether and where the Kalman filter estimated actual curve differs from the Kalman predicted curve.

Chapter 6

Continuous time state space modeling of panel data by means of SEM [1]

Abstract

Maximum likelihood parameter estimation of the continuous time linear stochastic SSM is considered on the basis of multiple subject discrete time data using a suitable SEM program. The exact discrete model (EDM) is employed which links the discrete time model parameters to the underlying continuous time model parameters by means of complex nonlinear restrictions. The SEM procedure covers the stationary as well as the nonstationary case. Nonstationarity is implemented by stepwise time-varying (piecewise time-invariant) parameter matrices and by parameter matrices varying continuously over time according to a polynomial scheme. The identification problem associated with continuous time state space modeling is treated and random subject effects are allowed to be part of the model.

6.1 Introduction

Over the past decades the study of change has become increasingly popular in the behavioral sciences. For the analysis of panel data SEM often is employed (Jöreskog, 1978; Jöreskog & Sörbom, 1985). It allows the specification of a dynamic causal structure in terms of latent variables underlying the discretely observed data. Although interest in the application of continuous time stochastic models is growing, SEM panel data analysis still typically relies on discrete time modeling.

By formulating the discrete time SEM model as a SSM (MacCallum & Ashby, 1986; Oud, van den Bercken, & Essers, 1990), the model can be made practically

---

[1] This article has been submitted: Oud, J.H.L., & Jansen, R.A.R.G. (1996). Continuous time state space modeling of panel data by means of SEM.

useful by employing techniques from control theory. Notable are the Kalman filter and Kalman smoother (Jazwinski, 1970), which allow the optimal estimation of the latent states or factor scores over time for individual subjects (Oud et al., 1990; Oud, van Leeuwe, & Jansen, 1993), as well as the application of optimal control procedures for controlling output behavior. The Kalman filter has been implemented in procedures for monitoring latent developmental characteristics of, for example, pupils and patients. The discrete time SSM covers a broad class of dynamic models. Examples are observed as well as latent ARMA and ARMAX models (the latter adding exogenous or input-variables to the ARMA model) of arbitrary order, longitudinal factor analysis models and state-trait models which include random subject effects. All can be estimated by means of a SEM program and can be used to estimate corresponding underlying continuous time versions.

Panel data, containing a large number N of independent replications of the entire vector of observed variables, allow the specification of a nonstationary model by means of time-varying parameters. This accommodates for gradual or sudden changes in developmental mechanisms. The option of time-invariance is retained in SEM state space modeling, however, by the possibility of specifying equality constraints between parameters.

Although, presumably because of the discrete time character of the data, continuous time models are not popular in behavioral science, several compelling reasons exist for using them. First, real life processes evolve in continuous time and are not restricted to the discrete observation time points the researcher happens to choose. Second, continuous time modeling provides the common base for an accurate comparison of differently spaced models of the same real life process. Furthermore, an effective solution of the missing data problem caused by different time sampling schemes for different subjects also requires consideration of continuous time modeling. Finally, the time points of the measurements in monitoring development by means of the Kalman filter or of the interventions for controlling output behavior, seldom coincide exactly with the time points of the discrete time model. In these cases too, it is important to be able to fill the gaps between the discrete time points of the model in order to use and process information at intermediate time points.

An overview of continuous time modeling in econometrics is given by Bergstrom (1988). In 1961-1962 Bergstrom (1988, p. 370) introduced the EDM which became central in many approaches to continuous time modeling. In social science deterministic differential equation models were first proposed by Coleman (1968). Oud (1978) adapted and used his procedure in structural equation modeling by means of the LISREL program (Jöreskog & Sörbom, 1989). Later, a systematic treatment of stochastic differential equation models was given by Tuma and Hannan (1984). Extending this approach, Arminger (1986) and Oud et al. (1993) employed the so-called 'indirect' method which consists of first estimating the discrete time parameters by means of a SEM program, and then in a second step deriving the continuous time parameter values using the EDM. The indirect method was heavily criticized by Hamerle, Nagl and Singer (1991). Because SEM programs like LISREL were not able to impose the nonlinear constraints between the continuous time and discrete time parameters in the EDM directly during estimation, they concluded that the "parameter estimates (which should not be computed!) are *misleading*, *useless* and *wrong*" (p. 212).

In the present article, we show that within the SEM approach maximum likelihood estimates of the continuous time parameters can be obtained by using the option of nonlinear constraints on parameters offered by recent versions of some SEM programs. We use the SEM program Mx (Neale, 1995), which in contrast to LISREL8 (Jöreskog & Sörbom, 1993) and most other programs utilizes matrix algebraic operations in the formulation of constraints. In addition to the time-invariant parameter matrices chosen by Singer (1993) and estimated by his program LSDE (Singer, 1991), our continuous time SSM also allows for different kinds of time-variance: piecewise time-invariant parameter matrices as well as parameter matrices varying continuously over time according to a polynomial scheme. These accommodate, respectively, for sudden and gradual changes in developmental mechanisms. We start with a discussion of the discrete time SSM. Then the EDM is derived from the linear stochastic differential equation. Next the special identification problems associated with continuous time state space modeling and the estimation procedure are explained. Finally, an educational research example is presented.

6.2 Discrete time SSM

The discrete time time-varying linear stochastic SSM consists of two equations: the dynamic part or state equation (Equation 6.1), which describes the dependence of the latent state variables in $\mathbf{x}_t$ on their lagged values in $\mathbf{x}_{t-1}$ and the static part or output equation, which connects the latent state variables to the observables in $\mathbf{y}_t$ (Equation 6.2):

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1} \quad \text{with} \quad cov(\mathbf{w}_{t-1}) = \mathbf{Q}_{t-1}\ , \tag{6.1}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \quad \text{with} \quad cov(\mathbf{v}_t) = \mathbf{R}_t\ . \tag{6.2}$$

The $m \times m$ matrix $\mathbf{A}_{t-1}$ in Equation 6.1 specifies the autoregressive and cross-lagged effects between the m state variables at successive discrete time points t and $t-1$: $t, t-1 \in \{t_0, t_0+1, \ldots, t_0+T-1\}$ for integers t_0 and $T \geq 2$, with t_0 the initial time point and T the total number of time points considered. The $p \times m$ matrix $\mathbf{C}_t$ in Equation 6.2 contains the loadings of the latent state variables on the p observed output variables in $\mathbf{y}_t$. Both equations allow for input-effects $\mathbf{B}_{t-1}\mathbf{u}_{t-1}$ and $\mathbf{D}_t\mathbf{u}_t$, respectively, from r fixed input-variables in the input-vectors $\mathbf{u}_{t-1}$ and $\mathbf{u}_t$.

The process errors in successive vectors $\mathbf{w}_t$ and the measurement errors in successive vectors $\mathbf{v}_t$ are assumed to have (a) zero expectations: $E(\mathbf{w}_t) = E(\mathbf{v}_t) = \mathbf{0}$ for all t, (b) zero covariances except, possibly, within vectors: $E(\mathbf{w}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$, $E(\mathbf{w}_t\ \mathbf{w}'_{t'}) = \mathbf{Q}_t\delta_{t-t'}$, $E(\mathbf{v}_t\ \mathbf{v}'_{t'}) = \mathbf{R}_t\delta_{t-t'}$ for all t and t' ($\delta_{t-t'}$ Kronecker's delta, being 0 if $t \neq t'$ and 1 if $t = t'$), (c) zero covariances with the initial state: $E(\mathbf{w}_t\ \mathbf{x}'_{t_0}) = E(\mathbf{v}_t\ \mathbf{x}'_{t_0}) = \mathbf{0}$ for all t. Finally, it is assumed that (d) the error vectors and the initial state are jointly multinormally distributed.

By specifying $\mathbf{B}_{t-1}\mathbf{u}_{t-1} = \mathbf{D}_t\mathbf{u}_t = \mathbf{0}$ and $E(\mathbf{x}_{t_0}) = E(\mathbf{y}_{t_0}) = \mathbf{0}$, implying $E(\mathbf{x}_t) = E(\mathbf{y}_t) = \mathbf{0}$ for all t, the SSM becomes zero-means first-order stationary. Under rather restrictive conditions, by specifying all four remaining matrices time-invariant ($\mathbf{A}, \mathbf{Q}, \mathbf{C}, \mathbf{R}$ instead of $\mathbf{A}_{t-1}, \mathbf{Q}_{t-1}, \mathbf{C}_t, \mathbf{R}_t$) and constraining $E(\mathbf{x}_t\ \mathbf{x}'_t) = E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0}) = \mathbf{\Phi}$ and thus $E(\mathbf{y}_t\ \mathbf{y}'_t) = E(\mathbf{y}_{t_0}\ \mathbf{y}'_{t_0}) = \mathbf{C\Phi C}' + \mathbf{R}$ to be equal for all t, the model becomes additionally second-order stationary.

First-order nonstationarity or nonconstant mean trajectories $E(\mathbf{x}_t)$ and $E(\mathbf{y}_t)$ result from the specification of input-variables and input-effects $\mathbf{B}_{t-1}\mathbf{u}_{t-1} \neq \mathbf{0}$ and $\mathbf{D}_t\mathbf{u}_t \neq \mathbf{0}$. In one case only a single, so-called unit input-variable is specified ($u_t = 1$ for all t), which is constant over time points as well as over subjects in the sample. Here the vectors $\mathbf{b}_{t-1}$ represent latent growth intercepts and the vectors $\mathbf{d}_t$ location parameters (origins) of the measurement scales. $\mathbf{C}_t$ relates the measurement scale units to those of the latent state variables. The model implies a mean trajectory which is common to all subjects in the sample. In another case the input-variables are all constant over time ($\mathbf{u}_t = \mathbf{u}_{t-k}$ for all t and $k > 0$) but, apart from the unit input-variable, varying over subjects (gender, socioeconomic status, etc.). Finally, additional input-variables may be specified that vary over time points as well as over subjects.

For nonconstant mean trajectories $E(\mathbf{x}_t)$ and $E(\mathbf{y}_t)$, the initial latent state means as well as the initial latent state variances may be chosen arbitrarily. Specifying

$$E(\mathbf{x}_{t_0}) = \mathbf{B}_{t_0-1}\mathbf{u}_{t_0-1}\ , \tag{6.3}$$

$$E(\mathbf{y}_{t_0}) = \mathbf{C}_{t_0}E(\mathbf{x}_{t_0}) + \mathbf{D}_{t_0}\mathbf{u}_{t_0}\ , \tag{6.4}$$

$E(\mathbf{x}_{t_0})$ is modeled by means of an additional matrix $\mathbf{B}_{t_0-1}$, to be specified entirely zero in case $E(\mathbf{x}_{t_0}) = \mathbf{0}$ but with just the elements corresponding to the unit input-variable in $\mathbf{u}_{t_0-1}$ free in case $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$. The value and identifiability of $E(\mathbf{x}_{t_0})$ depend on the choice of $\mathbf{D}_{t_0}$ as well as of the factor loading matrix $\mathbf{C}_{t_0}$. The latter additionally determines the value and identifiability of the initial state covariance matrix $\mathbf{\Phi}_{t_0} = E([\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})][\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]')$. In fact, these choices determine the measurement scales (origins and scale units) of the latent state variables. Special identification techniques are needed to guarantee that the latent scales maintain the same origins and measurement units across the whole time range (Oud et al., 1993, pp. 15-16).

In deriving the SEM model first write Equations 6.1-6.2 as follows:

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{t-1} & \mathbf{A}_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{w}_{t-1} \end{bmatrix}, \tag{6.5}$$

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}_t & \mathbf{C}_t \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_t \end{bmatrix}. \tag{6.6}$$

Collecting all input-variables in the input-vector $\mathbf{u}$ but specifying the constant (e.g. the unit input-variable) and other exactly linearly related input-variables only once in $\mathbf{u}$, and defining

$$\begin{array}{lll} \eta = [\mathbf{u}'\ \mathbf{x}']' & \text{with} & \mathbf{x} = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1} \cdots \mathbf{x}'_{t_0+T-1}]', \\ \zeta = [\mathbf{u}'\ \mathbf{w}']' & \text{with} & \mathbf{w} = [(\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0}))'\ \mathbf{w}'_{t_0} \cdots \mathbf{w}'_{t_0+T-2}]', \\ \mathbf{y} = [\mathbf{u}'\ \mathbf{y}'_0]' & \text{with} & \mathbf{y}_0 = [\mathbf{y}'_{t_0}\ \mathbf{y}'_{t_0+1} \cdots \mathbf{y}'_{t_0+T-1}]', \\ \varepsilon = [\mathbf{0}'\ \mathbf{v}']' & \text{with} & \mathbf{v} = [\mathbf{v}'_{t_0}\ \mathbf{v}'_{t_0+1} \cdots \mathbf{v}'_{t_0+T-1}]', \end{array} \tag{6.7}$$

the SEM model is derived as follows:

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & = & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \overline{\mathbf{B}} & \overline{\mathbf{A}} \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & + & \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} & , \\ \eta & = & B & \eta & + & \zeta & \end{array} \tag{6.8}$$

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} & = & \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \overline{\mathbf{D}} & \overline{\mathbf{C}} \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} & , \\ \mathbf{y} & = & \Lambda & \eta & + & \varepsilon & \end{array} \tag{6.9}$$

where all parameter matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, $\mathbf{C}_t$, $\mathbf{D}_t$ are put on the appropriate places in $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, $\overline{\mathbf{C}}$, $\overline{\mathbf{D}}$, respectively. Notice that in $\mathbf{x}$ the initial state $\mathbf{x}_{t_0}$ gets zero rows in $\overline{\mathbf{A}}$ but $\mathbf{B}_{t_0-1}$ in $\overline{\mathbf{B}}$ for modeling its mean $E(\mathbf{x}_{t_0})$, which therefore is subtracted from $\mathbf{x}_{t_0}$ in $\mathbf{w}$ (see Equation 6.2).

In maximum likelihood fitting of the discrete time SSM to the data in $\mathbf{Y}$ the SEM program minimizes

$$F_{ML} = \log \mid \Sigma \mid + \mathrm{tr}(\mathbf{S}\Sigma^{-1}) - \log \mid \mathbf{S} \mid - (p+q)\ , \tag{6.10}$$

for

$$\Sigma = E(\mathbf{y}\ \mathbf{y}') = \Lambda(\mathbf{I} - B)^{-1}\Psi(\mathbf{I} - B')^{-1}\Lambda' + \Theta\ , \tag{6.11}$$

where $\mathbf{\Psi} \equiv E(\boldsymbol{\zeta}\ \boldsymbol{\zeta}')$ and $\mathbf{\Theta} \equiv E(\boldsymbol{\varepsilon}\ \boldsymbol{\varepsilon}')$, on the basis of the sample moment matrix

$$\mathbf{S}_{(p+q)\times(p+q)} = \frac{1}{N}\mathbf{Y}\ \mathbf{Y}' = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} \mathbf{u}_i\ \mathbf{u}_i' & \frac{1}{N}\sum_{i=1}^{N} \mathbf{u}_i\ \mathbf{y}_{0i}' \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_{0i}\ \mathbf{u}_i' & \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_{0i}\ \mathbf{y}_{0i}' \end{bmatrix}, \tag{6.12}$$

where q is the number of (fixed) elements in $\mathbf{u}$. The set of distinct fixed values $\mathbf{u}_i$ in the sample must contain at least q linearly independent ones.

Adding to the state equation (Equation 6.1) random subject effects $\boldsymbol{\kappa}$

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \boldsymbol{\kappa} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1}\ , \tag{6.13}$$

leads to a model which became very popular in econometric panel analysis (e.g. Baltagi, 1995). Because of their constancy over time, the added normally distributed variables in $\boldsymbol{\kappa}$ are sometimes called 'trait' variables and the model a 'state-trait' model. A trait variable may be specified for each of the state variables. It can be characterized as a random (but constant over time) intercept term, to be contrasted to the fixed (but possibly time-varying) intercept associated with the unit input-variable. Because of the specification $E(\boldsymbol{\kappa}) = \mathbf{0}$, $\boldsymbol{\kappa}$ may be viewed as the subject specific deviation from the common fixed intercept. As the trait variable causes each subject's state to regress to or egress from the subject specific mean instead of the population mean, in a sense a subject specific $N = 1$ model is specified within the state-trait model for each of the subjects.

The state-trait model can be reformulated as a special case of the model in Equations 6.1-6.2

$$\begin{array}{ccccccccccc} \begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\kappa} \end{bmatrix} & = & \begin{bmatrix} \mathbf{A}_{t-1} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} & \begin{bmatrix} \mathbf{x}_{t-1} \\ \boldsymbol{\kappa} \end{bmatrix} & + & \begin{bmatrix} \mathbf{B}_{t-1} \\ \mathbf{0} \end{bmatrix} & \mathbf{u}_{t-1} & + & \begin{bmatrix} \mathbf{w}_{t-1} \\ \mathbf{0} \end{bmatrix} \\ \mathbf{x}^{\circ}{}_t & = & \mathbf{A}^{\circ}{}_{t-1} & \mathbf{x}^{\circ}{}_{t-1} & + & \mathbf{B}^{\circ}{}_{t-1} & \mathbf{u}_{t-1} & + & \mathbf{w}^{\circ}{}_{t-1} \end{array}, \tag{6.14}$$

$$\mathbf{y}_t = [\ \mathbf{C}_t\mathbf{0}\ ]\ \mathbf{x}^{\circ}{}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t\ , \tag{6.15}$$

and its parameters estimated by means of a SEM program. The constancy over time makes that $\boldsymbol{\kappa}$ should also be considered part of the initial state $\mathbf{x}_{t_0}$, so that $\boldsymbol{\kappa}$ and $\mathbf{x}_{t_0}$ cannot be assumed to be uncorrelated. The initial state covariance matrix in the SEM model becomes:

$$\mathbf{\Phi}^{\circ}{}_{t_0} = \begin{bmatrix} \mathbf{\Phi}_{x_{t_0}} & \mathbf{\Phi}_{x_{t_0},\kappa} \\ \mathbf{\Phi}_{\kappa,x_{t_0}} & \mathbf{\Phi}_{\kappa} \end{bmatrix}. \tag{6.16}$$

Significance tests on the existence of constant random subject effects can easily be performed: both the variances of the trait variables in $\mathbf{\Phi}_{\kappa}$ and their covariances with the initial state variables in $\Phi_{x_{t_0},\kappa}$ are testable quantities, required to be different from $\mathbf{0}$ (Baltagi, 1995, p. 125). Being constant variables, the trait variables need not be repeated for successive time points in the SEM model but may be specified once, while trait variables with nonsignificant variances may be deleted.

6.3 Continuous time SSM

Preparing the replacement of the discrete time state equation (Equation 6.13) by its continuous time analogue, we first give two reformulations of Equation 6.13. The first one changes nothing except the time scale:

$$\mathbf{x}_t = \mathbf{A}_{t-\Delta t}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \mathbf{B}_{t-\Delta t}\mathbf{u}_{t-\Delta t} + \mathbf{w}_{t-\Delta t} \ . \tag{6.17}$$

By means of Equation 6.17 researchers working with different time intervals between their measurements have the opportunity to put their interval on a common time scale as, for example, years, instead of each giving a different interpretation to interval 1. By labeling the parameter matrices with the appropriate Δt, it becomes immediately clear that an autoregression coefficient of .30, found over a period with $\Delta t = 1$ (a whole year), implies a stronger autoregression than the same coefficient found with only $\Delta t = .75$ (three quarters of a year). The second reformulation:

$$\frac{\Delta \mathbf{x}_t}{\Delta t} = \mathbf{A}^*_{t-\Delta t}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa}^* + \mathbf{B}^*_{t-\Delta t}\mathbf{u}_{t-\Delta t} + \mathbf{G}^*_{t-\Delta t}\frac{\Delta \mathbf{w}^*_t}{\Delta t} \ , \tag{6.18}$$

is in terms of $\frac{\Delta \mathbf{x}_t}{\Delta t} = \frac{\mathbf{x}_t - \mathbf{x}_{t-\Delta t}}{\Delta t}$, the average increase over the interval, and has $\mathbf{A}^*_{t-\Delta t} = \Delta t^{-1}(\mathbf{A}_{t-\Delta t} - \mathbf{I})$, $\boldsymbol{\kappa}^* = \Delta t^{-1}\boldsymbol{\kappa}$, $\mathbf{B}^*_{t-\Delta t} = \Delta t^{-1}\mathbf{B}_{t-\Delta t}$, $\mathbf{w}_{t-\Delta t} = \mathbf{G}^*_{t-\Delta t}\Delta\mathbf{w}^*_t$ where $\mathbf{w}^*_t$ is the standard discrete time random walk process. It enables a direct comparison of the results over different time intervals, although still in the form of a linear approximation of the underlying continuous time processes.

In continuous time $\mathbf{w}^*_t$ is replaced by the famous Wiener process $\mathbf{W}(t)$ or limiting form of the discrete time random walk process (Jazwinski, 1970, pp. 70-74) and Equation 6.18 by stochastic differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \gamma + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{G}(t)\frac{d\mathbf{W}(t)}{dt} \ . \tag{6.19}$$

$\mathbf{W}(t)$ is defined by having (Arnold, 1974, p. 46) stationary independent normally distributed increments $\mathbf{W}(t) - \mathbf{W}(s)$ over nonoverlapping intervals with $E[\mathbf{W}(t) - \mathbf{W}(s)] = \mathbf{0}$ and $cov[\mathbf{W}(t) - \mathbf{W}(s)] = |t-s|\mathbf{I}$ for all $s \neq t$ and $\mathbf{W}(0) = \mathbf{0}$ a.s. In contrast to the increments, $\mathbf{W}(t)$ is itself nonstationary: $E[\mathbf{W}(t)\ \mathbf{W}'(s)] =$

$\min(t,s)\mathbf{I}$ and thus $E[\mathbf{W}(t)\ \mathbf{W}'(t)] = t\mathbf{I}$. Whereas the variance of the increments depends on $|t-s|$ only, the variance of the elements of $\mathbf{W}(t)$ depends on t.

Although the sample trajectories of $\mathbf{W}(t)$ are continuous a.s., the independence of the increments over arbitrarily small intervals and the resulting unbounded variation (infinite length of the trajectory) over any finite interval in $[0,\infty)$ (Arnold, 1974, p. 48), makes them nowhere differentiable in the traditional sense. $\boldsymbol{\xi}(t) = \frac{d\mathbf{W}(t)}{dt}$ or the 'white noise' process in continuous time is therefore not a stochastic process in the ordinary sense. Used in many fields as a mathematical idealization of a rapidly fluctuating force, whose autocorrelation between instants t and s goes to zero rapidly for increasing $|t-s|$, $\boldsymbol{\xi}(t)$ exists within the theory of 'generalized functions' as a 'generalized stochastic process' (see Arnold, 1974, pp. 50-57; Ruymgaard & Soong, 1985, p. 85). It is the derivative of the Wiener process in this generalized function sense, having mean and covariance matrix $E[\boldsymbol{\xi}(t)] = \mathbf{0}$ and $E[\boldsymbol{\xi}(t)\ \boldsymbol{\xi}'(s)] = \mathbf{I}\delta(t-s)$, in terms of Dirac's delta function

$$\delta(t-s) = \begin{cases} 0 & \text{for } t \neq s \\ \infty & \text{for } t = s \end{cases} \qquad \int_{-\infty}^{\infty} \delta(t-s)ds = 1(t) = 1 \ .$$

A delta correlated force, occurring at isolated instant $t' = t - s = 0$ only, must be, in a certain sense, of infinite strength or variance to result in a unit integrated area. If $\delta(t-s)$ is a narrow impulse of unit area, $\mathbf{I}\delta(t-s)$ its matrix valued analogue, the sifting property of the delta function results in $\int_{-\infty}^{\infty} E[\boldsymbol{\xi}(t)\ \boldsymbol{\xi}'(s)]ds = \mathbf{I}$ and for nonstochastic $\mathbf{G}(t)$ in

$$\int_{-\infty}^{\infty} E[\mathbf{G}(t)\boldsymbol{\xi}(t)\ \boldsymbol{\xi}'(s)\mathbf{G}'(s)]ds = \int_{-\infty}^{\infty} \mathbf{G}(s)\mathbf{G}'(s)\delta(t-s)ds = \mathbf{G}(t)\mathbf{G}'(t) \ .$$

So, while the standard white noise impulse has area $\mathbf{I}$, its nonstandard version has area $\mathbf{G}(t)\mathbf{G}'(t)$.

Rewriting Equation 6.19 as

$$d\mathbf{x}(t) = [\mathbf{A}(t)\mathbf{x}(t) + \gamma + \mathbf{B}(t)\mathbf{u}(t)]dt + \mathbf{G}(t)d\mathbf{W}(t) \ , \tag{6.20}$$

and next in stochastic integral form as

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} [\mathbf{A}(s)\mathbf{x}(s) + \gamma + \mathbf{B}(s)\mathbf{u}(s)]ds + \int_{t_0}^{t} \mathbf{G}(s)d\mathbf{W}(s) \ , \tag{6.21}$$

the problem encountered in the definition of white noise $\boldsymbol{\xi}(t)$ shows up again in the second integral at the right hand side. The sample trajectories of the Wiener process being of unbounded variation prevents its definition as an ordinary Riemann-Stieltjes integral. However, $\int_{t_0}^{t} \mathbf{G}(s)d\mathbf{W}(s)$ is generally defined and solvable as an Itô stochastic integral with own rules of integration. In case $\mathbf{G}(t)$ is nonstochastic, the Itô integral coincides with the Wiener integral and can be solved by using the formula for integration by parts (Arnold, 1974; Ruymgaard & Soong, 1985), resulting in

$$E\{\int_{t_0}^{t}\mathbf{G}(s)d\mathbf{W}(s)[\int_{t_0}^{t}\mathbf{G}(u)d\mathbf{W}(u)]'\} = \int_{t_0}^{t}\mathbf{G}(s)\mathbf{G}'(s)ds\ , \tag{6.22}$$

and replacing $\mathbf{G}(t)$ by $\mathbf{\Phi}(t',t)\mathbf{G}(t)$, used in the next section, in (Arnold, 1974, p. 131)

$$\begin{aligned}&E\{\int_{t_0}^{t}\mathbf{\Phi}(t,s)\mathbf{G}(s)d\mathbf{W}(s)[\int_{t_0}^{t}\mathbf{\Phi}(t,u)\mathbf{G}(u)d\mathbf{W}(u)]'\} = \\ &\int_{t_0}^{t}\mathbf{\Phi}(t,s)\mathbf{G}(s)\mathbf{G}'(s)\mathbf{\Phi}'(t,s)ds\ .\end{aligned} \tag{6.23}$$

6.4 Deriving the EDM

The EDM relates the discrete time SSM parameters exactly to the underlying continuous time parameters. It is derived under much more general conditions for the continuous time parameter matrices than usually found in the literature (see e.g. Bergstrom, 1984, p. 1168; Hamerle et al., 1991, p. 214). In particular, $\mathbf{B}(t)$ and $\mathbf{G}(t)$ are allowed to vary continuously over the integration interval according to a general polynomial scheme. Furthermore, the input $\mathbf{u}(t)$ is allowed to vary linearly over the interval instead of being constant. To obtain an analytic solution, $\mathbf{A}(t)$ is assumed to be constant over the interval. As explained in the next section, however, it may vary over different intervals (stepwise time-varying or piecewise time-invariant). Thus,

$$\begin{aligned}\mathbf{A}(t) &= \mathbf{A}\ ,\\ \mathbf{B}(t) &= \sum_{i=0}^{n_b}(t-t_0)^i\mathbf{B}_i\ ,\\ \mathbf{G}(t) &= \sum_{i=0}^{n_g}(t-t_0)^i\mathbf{G}_i\ ,\\ \mathbf{u}(t) &= \mathbf{u}(t_0)+\mathbf{b}_{\mathbf{u}_{(t_0,t]}}(t-t_0) \quad \text{with} \quad \mathbf{b}_{\mathbf{u}_{(t_0,t]}} = \frac{\mathbf{u}(t)-\mathbf{u}(t_0)}{t-t_0}\ .\end{aligned} \tag{6.24}$$

The explicit solution of the stochastic integral Equation 6.21 is given by (Arnold, 1974, pp. 128-134; Ruymgaard & Soong, 1985, pp. 80-99)

$$\begin{aligned}\mathbf{x}(t) &= \mathbf{\Phi}(t,t_0)\mathbf{x}(t_0)+\int_{t_0}^{t}\mathbf{\Phi}(t,s)\gamma ds+\int_{t_0}^{t}\mathbf{\Phi}(t,s)\mathbf{B}(s)\mathbf{u}(s)ds+\\ &\quad\int_{t_0}^{t}\mathbf{\Phi}(t,s)\mathbf{G}(s)d\mathbf{W}(s)\ ,\end{aligned} \tag{6.25}$$

in terms of the state transition matrix $\Phi(t, t_0)$, and next, introducing the conditions of Equation 6.24, in particular $\Phi(t, t_0) = \exp[\int_{t_0}^{t} \mathbf{A}(s)ds] = e^{\mathbf{A}(t-t_0)}$, such that $\int_{t_0}^{t} \mathbf{A}(s)ds$ and $\mathbf{A}(t)$ commute (Zadeh & Desoer, 1963, p. 340), by

$$\begin{aligned} \mathbf{x}(t) &= e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^{t} e^{\mathbf{A}(t-s)} ds \gamma + \\ &\sum_{i=0}^{n_b} \int_{t_0}^{t} e^{\mathbf{A}(t-s)}(s-t_0)^i ds \mathbf{B}_i \mathbf{u}(t_0) + \\ &\sum_{i=0}^{n_b} \int_{t_0}^{t} e^{\mathbf{A}(t-s)}(s-t_0)^{i+1} ds \mathbf{B}_i \mathbf{b}_{\mathbf{u}_{(t_0,t]}} + \sum_{i=0}^{n_g} \tilde{\mathbf{w}}_{G_i,(t_0,t]} \\ \text{for} \qquad &\tilde{\mathbf{w}}_{G_i,(t_0,t]} = \int_{t_0}^{t} e^{\mathbf{A}(t-s)}(s-t_0)^i \mathbf{G}_i d\mathbf{W}(s) \ . \end{aligned} \tag{6.26}$$

Writing $\tilde{\mathbf{A}}_{t-t_0} \equiv e^{\mathbf{A}(t-t_0)}$ and

$$\tilde{\mathbf{A}}_{0,t-t_0} \equiv \int_{t_0}^{t} e^{\mathbf{A}(t-s)} ds = \mathbf{A}^{-1}[\tilde{\mathbf{A}}_{t-t_0} - \mathbf{I}] \ , \tag{6.27}$$

one finds for

$$\tilde{\mathbf{A}}_{i,t-t_0} \equiv \int_{t_0}^{t} e^{\mathbf{A}(t-s)}(s-t_0)^i ds = \int_{0}^{t-t_0} e^{\mathbf{A}(t-t_0-s)} s^i ds \ ,$$

in view of the recursive relation

$$\int e^{-\mathbf{A}s} s^i ds = -\mathbf{A}^{-1} e^{-\mathbf{A}s} s^i + i\mathbf{A}^{-1} \int e^{-\mathbf{A}s} s^{i-1} ds \ ,$$

the recursive relation

$$\tilde{\mathbf{A}}_{i,t-t_0} = i\mathbf{A}^{-1}[\tilde{\mathbf{A}}_{i-1,t-t_0} - i^{-1}(t-t_0)^i \mathbf{I}] \ , \tag{6.28}$$

for $i > 0$ and to be started with $\tilde{\mathbf{A}}_{0,t-t_0}$.

For $\tilde{\mathbf{w}}_{G_0,(t_0,t]}$ in the error term $\sum_{i=0}^{n_g} \tilde{\mathbf{w}}_{G_i,(t_0,t]}$ in Equation 6.26 one finds in view of Equation 6.23 and using $\text{row}(\mathbf{M}_1\mathbf{M}_2\mathbf{M}_3) = \mathbf{M}_1 \otimes \mathbf{M}_3' \text{row}\mathbf{M}_2$ ('row' the rowvec operation, putting the elements of a matrix rowwise in a column vector, 'irow' the inverse operation),

$$\begin{aligned} \mathbf{Q}_{0,t-t_0} \equiv cov(\tilde{\mathbf{w}}_{G_0,(t_0,t]}) &= \int_{t_0}^{t} e^{\mathbf{A}(t-s)} \mathbf{G}_0 \mathbf{G}_0' e^{\mathbf{A}'(t-s)} ds \\ &= \text{irow} \int_{t_0}^{t} e^{\mathbf{A}(t-s)} \otimes e^{\mathbf{A}(t-s)} \text{row}(\mathbf{G}_0\mathbf{G}_0') ds \\ &= \text{irow}[\int_{t_0}^{t} e^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})(t-s)} ds \text{row}(\mathbf{G}_0\mathbf{G}_0')] \ . \end{aligned}$$

Next, writing

$$\tilde{\tilde{\mathbf{A}}}_{t-t_0} \equiv e^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})(t-t_0)} , \tag{6.29}$$

and

$$\tilde{\tilde{\mathbf{A}}}_{0,t-t_0} \equiv \int_{t_0}^{t} e^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})(t-s)} ds = (\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})^{-1}[\tilde{\tilde{\mathbf{A}}}_{t-t_0} - \mathbf{I}\otimes\mathbf{I}] , \tag{6.30}$$

$\mathbf{Q}_{0,t-t_0}$ can be written as

$$\mathbf{Q}_{0,t-t_0} = \text{irow}[\tilde{\tilde{\mathbf{A}}}_{0,t-t_0}\text{row}(\mathbf{G}_0\mathbf{G}_0')] . \tag{6.31}$$

For arbitrary $\tilde{\mathbf{w}}_{G_i,(t_0,t]}$ in $\sum_{i=0}^{n_q} \tilde{\mathbf{w}}_{G_i,(t_0,t]}$

$$\begin{aligned}
\mathbf{Q}_{i,t-t_0} &\equiv cov(\tilde{\mathbf{w}}_{G_i,(t_0,t]}) \\
&= \int_{t_0}^{t} e^{\mathbf{A}(t-s)}(s-t_0)^i\mathbf{G}_i\mathbf{G}_i'(s-t_0)^i e^{\mathbf{A}'(t-s)} ds \\
&= \text{irow} \int_{t_0}^{t} e^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})(t-s)}(s-t_0)^{2i}\text{row}(\mathbf{G}_i\mathbf{G}_i')ds \\
&= \text{irow}[\tilde{\tilde{\mathbf{A}}}_{2i,t-t_0}\text{row}(\mathbf{G}_i\mathbf{G}_i')] ,
\end{aligned} \tag{6.32}$$

computed recursively from $\tilde{\tilde{\mathbf{A}}}_{0,t-t_0}$ by means of

$$\tilde{\tilde{\mathbf{A}}}_{j,t-t_0} \equiv j(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})^{-1}[\tilde{\tilde{\mathbf{A}}}_{j-1,t-t_0} - j^{-1}(t-t_0)^j\mathbf{I}\otimes\mathbf{I}] , \tag{6.33}$$

until $j = 2i$.

Fixing finally $t - t_0$ at the discrete time observation interval Δt the state equation of the EDM is derived

$$\mathbf{x}_t = \tilde{\mathbf{A}}_{\Delta t}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \sum_{i=0}^{n_b}[\tilde{\mathbf{A}}_{i,\Delta t}\mathbf{B}_i \ \ \tilde{\mathbf{A}}_{i+1,\Delta t}\mathbf{B}_i]\mathbf{u}^*_{t-\Delta t} + \mathbf{w}_{t-\Delta t} , \tag{6.34}$$

which is in the form of Equation 6.17 for extended input-vector

$$\mathbf{u}^*_{t-\Delta t} = \begin{bmatrix} \mathbf{u}_{t-\Delta t} \\ \mathbf{b}_{\mathbf{u}_{(t-\Delta t,t]}} \end{bmatrix} ,$$

and

$$\begin{aligned}
\mathbf{A}_{t-\Delta t} &= \tilde{\mathbf{A}}_{\Delta t} = e^{\mathbf{A}\Delta t} , \\
\boldsymbol{\Phi}_{\kappa} &= \tilde{\mathbf{A}}_{0,\Delta t}\boldsymbol{\Phi}_{\gamma}\tilde{\mathbf{A}}_{0,\Delta t}' \\
\text{with} \quad & \tilde{\mathbf{A}}_{0,\Delta t} = \mathbf{A}^{-1}[\tilde{\mathbf{A}}_{\Delta t} - \mathbf{I}] ,
\end{aligned}$$

$$
\begin{aligned}
\mathbf{B}_{t-\Delta t} &= \sum_{i=0}^{n_b}[\tilde{\mathbf{A}}_{i,\Delta t}\mathbf{B}_i \;\; \tilde{\mathbf{A}}_{i+1,\Delta t}\mathbf{B}_i] \\
\text{with} \quad & \tilde{\mathbf{A}}_{j,\Delta t} = j\mathbf{A}^{-1}[\tilde{\mathbf{A}}_{j-1,\Delta t} - j^{-1} \Delta t^j \mathbf{I}] \ , \\
\mathbf{Q}_{t-\Delta t} &= E(\mathbf{w}_{t-\Delta t}\mathbf{w}'_{t-\Delta t}) = \sum_{i=0,i'=0}^{n_g} \text{irow}[\tilde{\tilde{\mathbf{A}}}_{i+i',\Delta t}\text{row}(\mathbf{G}_i\mathbf{G}'_{i'})] \\
\text{with} \quad & \tilde{\tilde{\mathbf{A}}}_{\Delta t} = \mathrm{e}^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})\Delta t} \\
& \tilde{\tilde{\mathbf{A}}}_{0,\Delta t} = (\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})^{-1}[\tilde{\tilde{\mathbf{A}}}_{\Delta t} - \mathbf{I}\otimes\mathbf{I}] \\
& \tilde{\tilde{\mathbf{A}}}_{j,\Delta t} = j(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})^{-1}[\tilde{\tilde{\mathbf{A}}}_{j-1,\Delta t} - j^{-1}\Delta t^j\mathbf{I}\otimes\mathbf{I}] \ . \qquad (6.35)
\end{aligned}
$$

It should be noted that all the continuous time constraints imposed on the discrete time parameter matrices $\mathbf{A}_{t-\Delta t}$, $\mathbf{\Phi}_\kappa$, $\mathbf{B}_{t-\Delta t}$, and $\mathbf{Q}_{t-\Delta t}$ by Equation 6.35 can be expressed and applied during estimation by means of a SEM program like Mx.

Special cases of the EDM state equation 6.34 are given by Bergstrom (1984, p. 1168) and Hamerle, Singer and Nagl (1993, p. 292-293). For $\mathbf{B}(t) = \mathbf{B}_0$ and $\mathbf{G}(t) = \mathbf{G}_0$ Equation 6.34 becomes

$$
\begin{aligned}
\mathbf{x}_t &= \tilde{\mathbf{A}}_{\Delta t}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0\mathbf{u}_{t-\Delta t} + \\
& \quad \mathbf{A}^{-1}[\Delta t^{-1}\tilde{\mathbf{A}}_{0,\Delta t} - \mathbf{I}]\mathbf{B}_0(\mathbf{u}_t - \mathbf{u}_{t-\Delta t}) + \mathbf{w}_{t-\Delta t} \\
\text{with} \quad & \mathbf{Q}_{t-\Delta t} = \text{irow}[\tilde{\tilde{\mathbf{A}}}_{0,\Delta t}\text{row}(\mathbf{G}_0\mathbf{G}'_0)] \ , \qquad (6.36)
\end{aligned}
$$

where $\tilde{\mathbf{A}}_{\Delta t}$, $\mathbf{\Phi}_\kappa$, $\tilde{\mathbf{A}}_{0,\Delta t}$, and $\tilde{\tilde{\mathbf{A}}}_{0,\Delta t}$ are as in Equation 6.35. If additionally the input is assumed constant over each observation interval ($\mathbf{u}_t - \mathbf{u}_{t-\Delta t} = \mathbf{0}$) and thus taking the form of piecewise constant trajectories, the fourth term at the right hand side disappears.

6.5 Identification and estimation of the continuous time SSM parameters

The EDM in Equations 6.34-6.35 connects discrete time parameter matrices to continuous time analogues by complex nonlinear restrictions based on the matrix exponentials $\mathrm{e}^{\mathbf{A}\Delta t}$ and $\mathrm{e}^{(\mathbf{A}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{A})\Delta t}$. For evaluating the identification status of the continuous time parameters, it is assumed that the discrete time parameters are all identified, independently of the EDM restrictions. Preferably, in the first step of the proposed procedure, an independent discrete time SEM analysis actually is performed, yielding estimates of $6T-1$ parameter matrices: $\mathbf{A}_{t_i-\Delta t_i}, \mathbf{B}_{t_i-\Delta t_i}, \mathbf{Q}_{t_i-\Delta t_i}, \mathbf{C}_{t_i}, \mathbf{D}_{t_i}$, and $\mathbf{R}_{t_i}$ for each of the observation time points $t_1, \ldots, t_{T-1}$, and $E(\mathbf{x}^\circ{}_{t_0})$, $\mathbf{\Phi}^\circ{}_{t_0}$, $\mathbf{C}_{t_0}, \mathbf{D}_{t_0}$, and $\mathbf{R}_{t_0}$ for initial observation time point t_0. $E(\mathbf{x}^\circ{}_{t_0})$ combines $E(\mathbf{x}_{t_0})$ and $E(\boldsymbol{\kappa}) = \mathbf{0}$ (see Equation 6.14), while $\mathbf{\Phi}^\circ{}_{t_0}$ combines $\mathbf{\Phi}_{x_{t_0}}$, $\mathbf{\Phi}_\kappa$, and $\mathbf{\Phi}_{\kappa,x_{t_0}}$ (see Equation 6.16). It should be noted that the time sampling scheme may be irregular in the sense of observation intervals Δt_i of varying length between successive observation time points t_i.

As the next step in the procedure it is recommended to introduce in a new SEM analysis the continuous time parameter matrices and the associated nonlinear EDM restrictions according to Equation 6.35 in a piecewise time-invariant (stepwise time-varying) fashion, that is, applying the restrictions for each observation time point and associated observation interval separately, yielding estimates of $3T-1$ continuous time parameter matrices: a possibly distinct set $\mathbf{A}, \mathbf{B}_0, \mathbf{G}_0$, to be called $\mathbf{A}_{t_i}, \mathbf{B}_{0,t_i}, \mathbf{G}_{0,t_i}$, for each of the time points $t_1, \ldots, t_{T-1}$, and $E(\mathbf{x}^\circ{}_{t_0})$ and $\mathbf{\Phi}^\circ{}_{t_0}$, for initial observation time point t_0. $E(\mathbf{x}^\circ{}_{t_0})$ combines $E[\mathbf{x}(t_0)]$ and $E(\gamma) = \mathbf{0}$, while $\mathbf{\Phi}^\circ{}_{t_0}$ combines $\mathbf{\Phi}_{x(t_0)}$, $\mathbf{\Phi}_\gamma$, and $\mathbf{\Phi}_{\gamma,x(t_0)}$. To get $\mathbf{\Phi}_\gamma$ at the place of $\mathbf{\Phi}_\kappa$ in $\mathbf{\Phi}^\circ{}_{t_0}$ (see Equation 6.16), at each t_i the appropriate $\tilde{\mathbf{A}}_{0,\Delta t}$ (see Equation 6.35) instead of $\mathbf{I}$ has to be specified in Equation 6.14 (in the upper right hand corner of $\mathbf{A}^\circ{}_{t-1}$).

As the replacement of the $3T-1$ discrete time parameter matrices of the state equation by the $3T-1$ continuous time parameter matrices does not increase the number of distinct parameters, the necessary identification condition is fulfilled. Hamerle et al. (1993) have argued convincingly that sufficiency is not guaranteed, however. Calling for convenience arbitrary $\mathbf{A}_{t_i}$ $\mathbf{A}$ and assuming all eigenvalues of the real matrices $\mathbf{A}$ and $\mathbf{A}_{t-\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}$ distinct, $\mathbf{A}$ and $\mathbf{A}_{t-\Delta t}$ can be diagonalized: $\mathbf{A} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}$ and $\mathbf{A}_{t-\Delta t} = \mathbf{T}\mathrm{e}^{\mathbf{\Lambda}\Delta t}\mathbf{T}^{-1}$, respectively, with $\mathbf{T}$ the common matrix of eigenvectors and $\mathbf{\Lambda}$ the diagonal matrix of possibly complex eigenvalues $a \pm bj$ (j the imaginary number $\sqrt{-1}$). Writing $\mathbf{\Lambda} = \mathbf{\Lambda}_a + \mathbf{\Lambda}_b j$ with $\mathbf{\Lambda}_a$ containing the real part coefficients a and $\mathbf{\Lambda}_b$ the imaginary part coefficients $\pm b$ in conjugate pairs, it is easily seen that adding $\pm k 2\pi \bigtriangleup t^{-1}$ for arbitrary positive k to the conjugate pairs $\pm b$ in $\mathbf{\Lambda}_b$ does not change $\mathbf{A}_{t-\Delta t}$. Each k leads to a different $\mathbf{A}$ but to the same $\mathbf{A}_{t-\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}$ and thus an identification problem is implied (Hamerle et al., 1991; Phillips, 1973). This identification problem only arises in the case of complex eigenvalues, however, and, because complex eigenvalues show up in conjugate pairs for a real $\mathbf{A}$, only if the number of state variables m (number of rows or columns in $m \times m$ matrix $\mathbf{A}$) is greater than one. Complex eigenvalues imply oscillatory movements which in different frequencies may coincide at certain time points and therefore become observationally indistinguishable at those time points. In some fields oscillatory movements can be excluded a priori, so that the restriction $\mathbf{\Lambda}_b = \mathbf{0}$ would solve the problem.

Even when for $m > 1$ complex eigenvalues cannot be excluded, the problem is limited in the sense that there exists only a finite number of positive integers k for which $\mathbf{A}$ leads to a real $\mathbf{G}_0$ (positive definite $\mathbf{G}_0\mathbf{G}_0'$; Hansen & Sargent, 1983). The size of the finite set of observationally equivalent structures or 'aliases' additionally depends on the observation interval Δt, a smaller Δt leading to less aliases and a sufficiently small Δt to identification (Bergstrom, 1988, p. 379). One could inspect the set of aliases and reject theoretically improbable ones a posteriori. It can be argued, however, that aliases should be excluded a priori by imposing extra restrictions on the parameters in $\mathbf{A}$, $\mathbf{G}_0$ or other continuous time parameter matrices.

The identification problem is solved in an alternative way (Hamerle, et al., 1993, pp. 293-294), when instead of the frequent assumption of piecewise constant input trajectories: $\mathbf{u}_t = \mathbf{u}_{t-\Delta t}$ and thus $\mathbf{b}_{\mathbf{u}_{(t-\Delta t,t]}} = \mathbf{0}$ in Equation 6.24, the trajectories of some or all input-variables are approximated by piecewise linear functions: $\mathbf{b}_{\mathbf{u}_{(t-\Delta t,t]}} \neq \mathbf{0}$ and thus $\mathbf{u}_t - \mathbf{u}_{t-\Delta t} \neq \mathbf{0}$ in Equation 6.36, and when at the same time the submatrix $\mathbf{B}_0^*$ of $\mathbf{B}_0$ corresponding to those input-variables has column rank $\geq m-1$. The latter part of the condition is automatically satisfied for models with the number of state variables $m \leq 2$. This identification condition follows from Equation 6.36, showing that observationally equivalent $\mathbf{A}$ must have, in addition to equal $\tilde{\mathbf{A}}_{\Delta t}$ and $\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*$, also equal

$$\begin{aligned}
&\mathbf{A}^{-1}[\Delta t^{-1}\tilde{\mathbf{A}}_{0,\Delta t} - \mathbf{I}]\mathbf{B}_0^* = \\
&\Delta t^{-1}\mathbf{A}^{-1}\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^* - \mathbf{A}^{-1}[\tilde{\mathbf{A}}_{\Delta t} - \mathbf{I}]^{-1}[\tilde{\mathbf{A}}_{\Delta t} - \mathbf{I}]\mathbf{B}_0^* = \\
&\Delta t^{-1}\mathbf{A}^{-1}\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^* - [\tilde{\mathbf{A}}_{\Delta t} - \mathbf{I}]^{-1}\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^* \ ,
\end{aligned}$$

($\mathbf{A}^{-1}$ commutes with $[\tilde{\mathbf{A}}_{\Delta t} - \mathbf{I}]^{-1}$). Hence, for an observationally equivalent $\mathbf{A}_k$ of $\mathbf{A}$

$$\Delta t^{-1}\mathbf{A}_k^{-1}\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^* = \Delta t^{-1}\mathbf{A}^{-1}\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^* \ ,$$

or

$$\mathbf{A}_k^{-1}[\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*] = \mathbf{A}^{-1}[\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*] \ ,$$

or ($\mathbf{A}_k$ commutes with $\mathbf{A}$)

$$\mathbf{A}_k[\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*] = \mathbf{A}[\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*] \ .$$

If $\mathbf{B}_0^*$ and thus $[\tilde{\mathbf{A}}_{0,\Delta t}\mathbf{B}_0^*]$ has column rank $\geq m-1$, this can only be the case when $\mathbf{A}_k = \mathbf{A}$, as for $\mathbf{A}_k \neq \mathbf{A}$ $\mathbf{A}_k - \mathbf{A}$ has rank ≥ 2. Therefore $\mathbf{A}$ is identified.

As the next step in the procedure, after guaranteeing identification of the continuous time parameters, it is recommended to use again Equations 6.34-6.35 in a new run of the SEM program but introducing complete time-invariance over the entire time range considered by imposing the equality constraints $\mathbf{A}_{t_i} = \mathbf{A}, \mathbf{B}_{0,t_i} = \mathbf{B}_0, \mathbf{G}_{0,t_i} = \mathbf{G}_0$ for all t_i. If the fit of this SEM model is judged not to deteriorate in comparison to the stepwise time-varying model estimated in the previous step, a particularly easily interpretable and parsimonious continuous time model is found and the analysis could be finished. If, however, the fit is judged to deteriorate unacceptably, $\mathbf{B}(t)$ and $\mathbf{G}(t)$ could be made continuously time-varying by introducing successively higher order matrices $\mathbf{B}_i$ and $\mathbf{G}_i$ in Equations 6.24 and

6.34-6.35 until the fit becomes satisfactory. In comparison to the stepwise time-varying model, this offers not only the advantage of avoiding sudden parameter changes when not meaningful theoretically, but yields also a more parsimonious model, when the number of higher order matrices can be kept small. The model keeps being identified as long as n_b and n_g do not exceed $T - 1$ and the appropriate constraints are applied between the observation time points t_i. These extra constraints are explained for $\mathbf{B}(t)$ below but are to be applied analogously for $\mathbf{G}(t)$.

A linearly time-varying $\mathbf{B}(t)$ has between initial observation time point t_0 and the next observation time point t_1, for all $t_0 < t \leq t_1$,

$$\mathbf{B}(t) = \mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1}(t - t_0) \ .$$

Next, for all $t_1 < t \leq t_2$ in the second observation interval, going from t_1 to t_2,

$$\mathbf{B}(t) = \mathbf{B}_{0,t_2} + \mathbf{B}_{1,t_2}(t - t_1) \ ,$$

which for a continuously time-varying matrix $\mathbf{B}(t)$ over successive observation intervals according to the same linear scheme should be equal to

$$\begin{aligned}
\mathbf{B}(t) &= \mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1}(t - t_1 + t_1 - t_0) \\
&= [\mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1}(t_1 - t_0)] + \mathbf{B}_{1,t_1}(t - t_1) \\
&= [\mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1} \triangle t_1] + \mathbf{B}_{1,t_1}(t - t_1) \ .
\end{aligned}$$

This means that the constraints to be introduced in the SEM model for observation interval $\triangle t_2 = t_2 - t_1$ should read

$$\begin{aligned}
\mathbf{B}_{0,t_2} &= [\mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1} \triangle t_1] \\
\mathbf{B}_{1,t_2} &= \mathbf{B}_{1,t_1} \ ,
\end{aligned}$$

and generally for arbitrary observation interval $\triangle t_i = t_i - t_{i-1}$

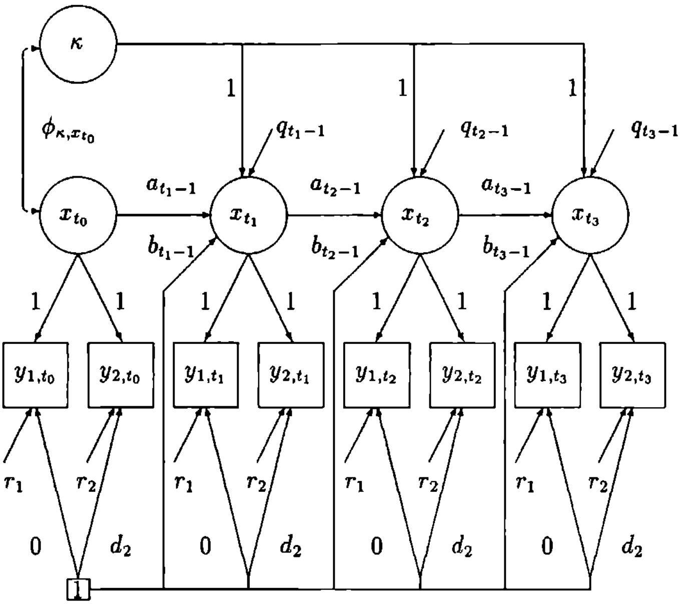
$$\begin{aligned}
\mathbf{B}_{0,t_i} &= [\mathbf{B}_{0,t_{i-1}} + \mathbf{B}_{1,t_{i-1}} \triangle t_{i-1}] \\
\mathbf{B}_{1,t_i} &= \mathbf{B}_{1,t_{i-1}} \ . \qquad (6.37)
\end{aligned}$$

Analogously, the constraints appropriate for higher order polynomial schemes are derived. Identification is a result of the constraints imposed on the parameter matrices of later observation intervals by the parameter matrices of preceding observation intervals. As all $\tilde{\mathbf{A}}_{j,\Delta t_i}$ are identified and invertible, and all $\mathbf{B}_{t_i - \Delta t_i} = \sum_{i'=0}^{n_b}[\tilde{\mathbf{A}}_{i',\Delta t_i}\mathbf{B}_{i',t_i} \ \tilde{\mathbf{A}}_{i'+1,\Delta t_i}\mathbf{B}_{i',t_i}]$ (see Equation 6.35) are identified, a general solvable linear system results. For the linear scheme ($n_b = 1$), for instance, identified $\mathbf{B}_{t_2 - \Delta t_2}$ and $\mathbf{B}_{t_1 - \Delta t_1}$ contain identified submatrices

$$\begin{aligned}
\mathbf{M}_{t_2 - \Delta t_2} &= \tilde{\mathbf{A}}_{0,\Delta t_2}\mathbf{B}_{0,t_2} + \tilde{\mathbf{A}}_{1,\Delta t_2}\mathbf{B}_{1,t_2} \\
&= \tilde{\mathbf{A}}_{0,\Delta t_2}\mathbf{B}_{0,t_1} + [\tilde{\mathbf{A}}_{0,\Delta t_2} \triangle t_1 + \tilde{\mathbf{A}}_{1,\Delta t_2}]\mathbf{B}_{1,t_1} \ , \\
\mathbf{M}_{t_1 - \Delta t_1} &= \tilde{\mathbf{A}}_{0,\Delta t_1}\mathbf{B}_{0,t_1} + \tilde{\mathbf{A}}_{1,\Delta t_1}\mathbf{B}_{1,t_1} \ ,
\end{aligned}$$

and in general these equations can be solved for $\mathbf{B}_{0,t_1}$ and $\mathbf{B}_{1,t_1}$, for $\tilde{\mathbf{A}}_{0,\Delta t_i}^{-1}\tilde{\mathbf{A}}_{1,\Delta t_i} = \mathbf{A}^{-1} - \Delta t_i[\tilde{\mathbf{A}}_{\Delta t_i} - \mathbf{I}]^{-1}$ and $\tilde{\mathbf{A}}_{0,\Delta t_2}^{-1}\mathbf{M}_{t_2-\Delta t_2} - \tilde{\mathbf{A}}_{0,\Delta t_1}^{-1}\mathbf{M}_{t_1-\Delta t_1} = (\Delta t_1\mathbf{I} + \Delta t_1[\tilde{\mathbf{A}}_{\Delta t_1} - \mathbf{I}]^{-1} - \Delta t_2[\tilde{\mathbf{A}}_{\Delta t_1} - \mathbf{I}]^{-1})\mathbf{B}_{1,t_1}$. Thus $\mathbf{B}(t) = \mathbf{B}_{0,t_1} + \mathbf{B}_{1,t_1}(t - t_0)$ is automatically identified for $\Delta t_1 = \Delta t_2$. For special cases, when $\Delta t_1 \neq \Delta t_2$, extra restrictions could be necessary, the simples ones being inequality constraints $\Delta t_1 + \Delta t_1(\mathrm{e}^{\lambda_i \Delta t_1} - 1)^{-1} - \Delta t_2(\mathrm{e}^{\lambda_i \Delta t_2} - 1)^{-1} \neq 0$ with respect to all distinct eigenvalues λ_i of $\mathbf{A}$.

Fig. 6.1: Discrete time SEM model for decoding speed.



6.6 Educational research example

At each of four half year distant time points two forms of a decoding speed test, were taken from 740 primary school pupils during first and second grade. The purpose of the study was estimating and evaluating the pupils' progress in decoding speed on a continuous time scale. First a discrete time SSM (Equation 6.14 or 6.17 and Equation 6.15) was specified and put in SEM form. Figure 6.1 shows the SEM model. The circles represent the latent decoding speed or state variables x_{t_i}. The time points $t_0 = t_1 - 1, t_1 = t_2 - 1, t_2 = t_3 - 1, t_3$ are separated by a constant unit time interval $\Delta t_i = 1$, corresponding to the constant half year period between the measurements. First-order nonstationarity (changing latent and observed means) is realized by the specification of the unit input-variable, shown in the small square and defining intercepts in the state and output equations. In addition to the autoregressive parameters a_{t_i-1}, latent growth intercepts b_{t_i-1}, and process error variances q_{t_i-1}, which are all time-varying parameters, the state equation also contains the trait variable κ, which for each subject has a constant value over time. Its variance and covariances with the state variables are brought into the model by specifying ϕ_κ and $\phi_{\kappa,x_{t_0}}$, respectively, in matrix $\mathbf{\Phi}^\circ{}_{t_0}$ (see Equation 6.16).

For disattenuating the state equation parameter estimates for measurement errors in the observed variables y_{1,t_i} and y_{2,t_i}, an output equation or measurement model was specified as part of the SEM model. All the measurement model parameters were chosen time-invariant, making the latent measurement scale characteristics constant over time. Because the two forms of the decoding speed test had the property of so-called 'essential tau-equivalence', the factorloadings c_1 and c_2, representing the effects of the latent variables on the observed variables, were both fixed to unity but the measurement equation intercepts d_1 and d_2 as well as the measurement error variances r_1 and r_2 were allowed to differ between both instruments. The choices of $c_1 = 1$ and $d_1 = 0$ for the first instrument cause the scale unit and origin of the latent measurement scale to become equal to those of the first instrument.

The total number of distinct free parameters in the SEM model is 17, which for 9 observed variables (8 observed decoding speed variables plus the unit input-variable) leaves 28 degrees of freedom for the model as a whole. Not mentioned in Figure 6.1 are the initial state mean $E(x_{t_0})$, initial state variance $\phi_{x_{t_0}}$, trait variance ϕ_κ, and moment 1 of the unit input-variable. Estimating the latter, whose estimate should be exactly 1, is one of the checks on the correctness of the maximum likelihood estimate.

By means of the SEM program Mx we estimated several models, starting with the time-varying discrete time model (DM) explained above. The parameter estimates of the DM and other models are shown in Table 6.1. Each time point in Table 6.1 indicates the end point of the observation interval the parameter refers to. The estimates of the time-varying autoregressive parameters in the DM show a moderate decrease over time, those of the intercepts increase, especially from t_2 to t_3, and those of the unexplained variances first increase and then decrease. In view of the standard errors, all parameter estimates in Table 6.1, including those for the trait variance and state-trait covariance, are highly significant, for the DM as well as for the other models. The parameter estimates of the measurement model, that is, the estimates of d_2, r_1 and r_2 in Figure 6.1, are not present in Table 6.1. These parameter estimates and their standard errors showed only minor differences between the models. The d_2 estimates all equaled -1.075 with a standard error of 0.082, meaning that the second test form is more difficult than the first one. Those for r_1 ranged from 10.056 to 10.171 with standard errors from 0.476 to

0.495, whereas the estimates of r_2 ranged from 9.604 to 9.821 with standard errors from 0.474 to 0.488.

Next, in accordance with the procedure recommended above, the continuous time parameters and associated nonlinear restrictions were introduced into the DM in a piecewise time-invariant fashion, leading to the EDM with continuous time parameter estimates given in the fourth column of Table 6.1. As the EDM replaces the DM discrete time parameters by continuous time parameters in a one-to-one way, the degrees of freedom and model fit values do not differ from those of the DM.

Next, complete time-invariance was introduced into the EDM by specifying all corresponding continuous time parameters to be equal over time. So, the sudden jumps implied by piecewise time-invariance (stepwise time-variance) were avoided by forcing strict constancy over time. This model, however, called EDM0 in Table 6.1, led to a serious fit deterioration (χ^2 increases from 120.54 to 295.50, while the increase in *df* is only 6; also the *AIC* (Akaike, 1974) shows a large increase). The most conspicuous difference with the EDM is the much higher drift parameter value (less negative feedback), which clearly is compensated by a much smaller trait variance.

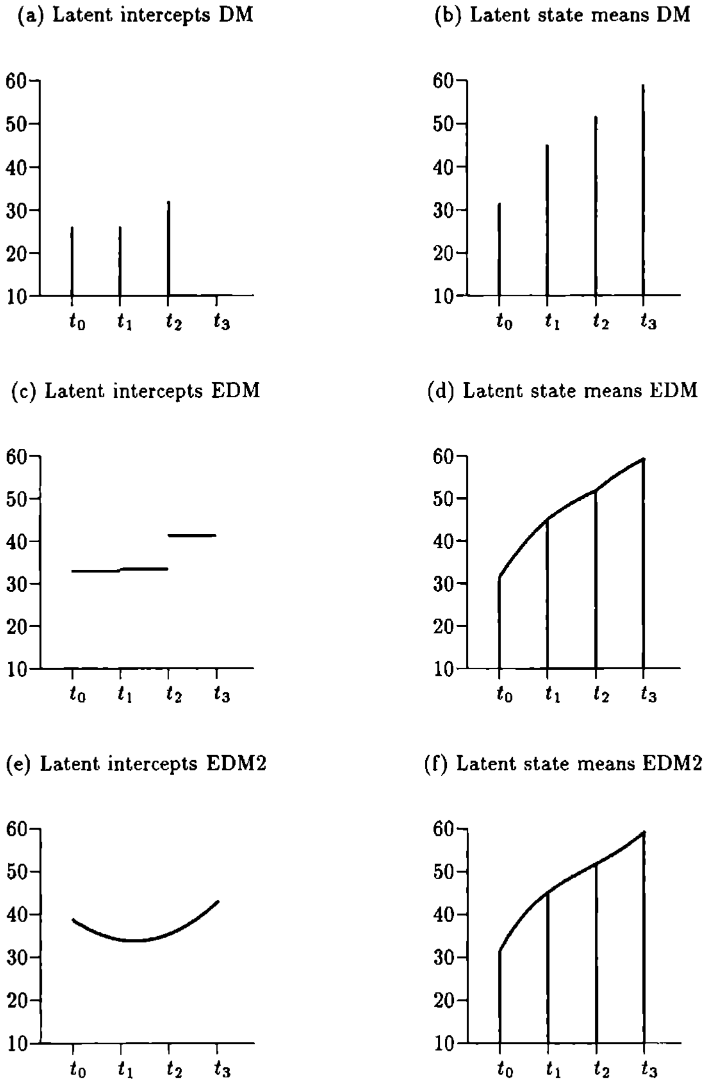
To improve model fit, we made the model time-varying again but now continuously instead of stepwise and as parsimoniously as possible. Because of the relatively large difference in the estimated latent intercept values over time in the EDM, we decided to concentrate on the intercept only, keeping other model parameters time-invariant. First a single linear component was added to the intercept, $b(t) = b_0 + b_1(t - t_0)$, leading to model EDM1. This did not realize an important model fit improvement, however. It had the opposite effect on the drift parameter value and trait variance in comparison to the EDM: a much lower drift parameter value of -.925 (very high negative feedback) and a much higher trait variance of 152.80, explaining almost all state variance. Finally, we added a quadratic component to the intercept, $b(t) = b_0 + b_1(t - t_0) + b_2(t - t_0)^2$. This model, called EMD2 in Table 6.1, was judged to reproduce the observed moment matrix relatively well. With 1 *df* less than in the EDM1, it realizes a χ^2 drop of 92. In addition, the drift parameter and trait variance regained the order of magnitude found in the EDM.

To illustrate some of the results with regard to modeling latent intercepts in discrete and continuous time and its effects on the latent means, we plotted in Figure 6.2 the latent intercepts and latent means of three relatively well-fitting models in Table 6.1: time-varying discrete time model DM, piecewise time-invariant model EDM, and model EDM2 with the quadratically time-varying intercept. Figure 6.2a shows the intercepts of the DM. At t_0 the intercept b_{t_1-1} of the first time interval is shown. This influences additively the latent mean value at time point t_1 in Figure 6.2b (cfr. Equation 6.17), the same is true for the intercepts b_{t_2-1} and b_{t_3-1} at time points t_1 and t_2, influencing additively the latent mean values at time points t_2 and t_3, respectively. In Figure 6.2c the continuous time intercept values are plotted for the piecewise time-invariant model EDM. Here the same pattern arises as for the discrete time intercepts of the DM, both sets showing comparable differences between the values at t_0, t_1, and t_2. However, the continuously and rather smoothly changing latent means in Figure 6.2d arise as complex nonlinear

Tab. 6.1: SEM solutions of the latent parts of the discrete time model (DM) and four continuous time exact discrete models (EDM, EDM0, EDM1, EDM2); standard errors are indicated between parentheses.

		DM	EDM	EDM0		EDM1		EDM2
Autoregressive	t_1	0.604	-0.503	-0.391		-0.925		-0.563
or drift		(0.047)	(0.079)	(0.018)		(0.084)		(0.08)
parameters	t_2	0.567	-0.552					
		(0.050)	(0.083)					
	t_3	0.524	-0.609					
		(0.049)	(0.085)					
Intercepts	t_1	25.95	32.96	27.63	b_0	46.18	b_0	38.42
		(1.49)	(3.14)	(0.88)		(3.04)		(2.79)
	t_2	26.11	33.42		b_1	4.63	b_1	-7.71
		(2.25)	(4.09)			(0.70)		(1.26)
	t_3	31.92	41.24				b_2	3.03
		(2.53)	(4.78)					(0.29)
Unexplained	t_1	11.08	17.46	22.18		26.17		20.32
variances		(1.66)	(2.27)	(1.37)		(1.89)		(1.44)
	t_2	14.51	24.39					
		(1.65)	(2.80)					
	t_3	9.34	16.71					
		(1.18)	(2.53)					
Initial state	t_0	31.51	31.51	31.59		31.56		31.51
mean		(0.51)	(0.51)	(0.51)		(0.51)		(0.51)
State means	t_1	44.97	44.97	44.25		44.37		44.97
	t_2	51.59	51.59	52.82		52.47		51.59
	t_3	58.96	58.96	58.62		58.70		58.96
Initial state		184.94	184.93	186.60		186.88		187.00
variance		(9.88)	(9.88)	(9.95)		(9.98)		(10.04)
Trait variance		35.05	56.54	27.38		152.80		57.83
		(7.58)	(16.41)	(3.14)		(28.92)		(16.43)
Initial state-		64.38	81.57	54.93		144.70		84.08
trait covariance		(9.09)	(14.51)	(4.71)		(16.39)		(14.35)
χ^2		120.54	120.54	295.50		244.62		152.17
df		28	28	34		33		32
AIC		64.54	64.54	227.50		178.62		88.17

Fig. 6.2: Latent intercepts and latent state means in the DM, EDM, and EDM2.



functions of the piecewise time-invariant intercept values and the completely time-invariant drift parameter (using Equation 6.34 for increasing Δt). Nevertheless, the sudden intercept value changes in Figure 6.2c, especially at t_2, are still clearly visible in the resulting means curve in Figure 6.2d. In Figure 6.2e the quadratically changing intercepts are plotted for the EDM2. While the EDM2 intercept values are comparable in magnitude to those of the EDM at t_1, t_2, and t_3, the rather artificial discrete time lag inherent in discrete time modeling makes that the EDM2 values at t_1, t_2, and t_3 should be compared to the DM values at lagged time points t_0, t_1, and t_2. In contrast to the DM and the EDM both the intercepts and means (Figure 6.2f) change as perfectly smooth, continuous functions of time.

Finally, in Figure 6.3, on the basis of EDM2 the Kalman filter and smoother estimates are plotted of the latent state and trait trajectories of one individual pupil in the sample. While the updates at observation time points t_1, t_2, and t_3 make the filter trajectories discontinuous and diminish the confidence interval widths at those time points, the backward recursion of the smoother, which starts at the last filter update at t_3, eliminates the discontinuities. By definition the trait trajectory is a constant curve (with zero mean in the population), but the filter is seen to improve the estimate of this constant value at successive time points. The smoother confidence interval bands show that the trait value of this particular subject is significantly below population mean zero and also its state is everywhere significantly below the population mean (cfr. Figure 6.2f). On the basis of continuous time modeling this conclusion is made validly for the entire continuous time range and need not be restricted to the discrete observation time points only.

6.7 Discussion

While for more than three centuries continuous time modeling by means of differential equations is the standard approach of dynamic phenomena in the physical sciences, in social and behavioral science continuous time modeling is still rare. The reasons are, most probably, that the theory of stochastic differential equations has been developed mainly for application in specific fields of physics and that the nonlinear constraints needed for estimating the parameters of stochastic differential equations on the basis of discrete time data were not generally available in statistical application software. In this article it is shown that an appropriate SEM program can be used to do the job on the basis of discrete time panel data. We illustrated the procedure by means of a simple unidimensional SSM where the intercept is modeled to influence the individual subject's state and population's state mean in the form of a continuous process, time-invariant as well as in different ways time-varying.

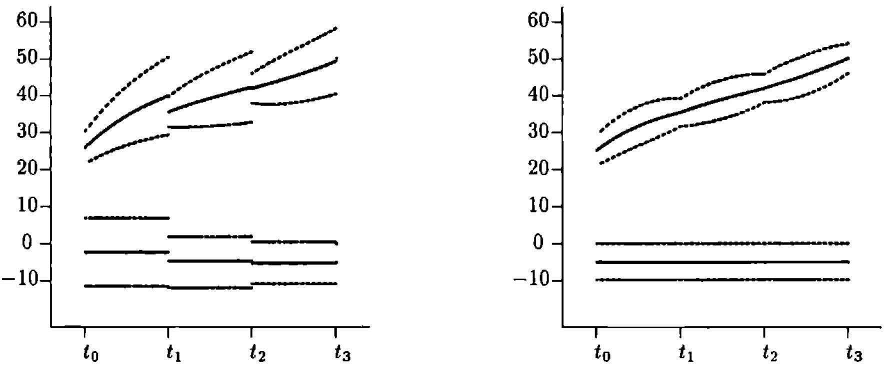
The proposed SEM procedure allows also missing data and irregular time sampling schemes to be processed. For SEM estimation purposes, the EDM can be considered just a special case of the discrete time SSM. Therefore, in the case of missing data the same EM procedure is applicable as in the discrete time case

(Jansen & Oud, 1995) The EM procedure can also easily be extended to irregular time sampling for different subjects, as due to continuous time modeling conditional expectations (Kalman smoother values) can be computed at arbitrary time points and thus for all subjects at identical time points

A topic for further research concerns the drift matrix $\mathbf{A}(t)$ We showed how the matrices $\mathbf{B}(t)$ and $\mathbf{G}(t)$ of the stochastic differential equation (Equation 6 19) can be modeled in continuous time in a time varying way, namely by defining and solving the relevant integrals of (Equation 6 25) in terms of a polynomial scheme We did not handle the drift matrix $\mathbf{A}(t)$ in this way, because introducing the drift matrix as a function of polynomial terms yields integrals which cannot be solved analytically It is worthwhile, however, to look for an appropriate numerical solution In the behavioral sciences, the hypothesis of autoregression effects varying over time, from observation period to observation period, seems a plausible one and is often entertained in discrete time research projects It is important to be able to generalize this hypothesis to continuous time by means of continuously time varying parameters

Fig 6 3 EDM2 Kalman filter and smoother estimates of one pupil's state (upper solid curve) and trait (lower solid curve) trajectories with 95% *confidence interval bands (dashed curves)*

(a) EDM2 Kalman filter estimates (b) EDM2 Kalman smoother estimates



Chapter 7

Conclusions, implications for monitoring and future research

In chapters 2-6 a number of research topics in SEM state space modeling of panel data have been discussed. Each contribution, in fact, illustrates the use of the SEM state space approach in the analysis of behavioral science data. Both SEM and the state space model can be employed for the modeling of dynamic stochastic processes and both cover a wide range of dynamic systems. The SEM approach provides information on the (sub)population level, and is an important means for model construction and testing, model selection and parameter estimation. The state space model model allows the application of the Kalman filter and Kalman smoother as optimal state estimators. These estimators combine the population information in the form of the estimated state space model parameters with subject specific observations. It is shown that the combined approach of SEM and the state space model is useful for the construction of monitoring systems on the basis of repeated measurements data. Below the relevance of each chapter as regards the construction of monitoring systems within primary school education is stressed. The major findings are restated as well as a number of topics that have implications for future research.

Several findings of the introductory chapter are important for the construction of monitoring systems in primary school education. In practice monitoring can be very diverse. Especially in the early phase practical aspects and insights guide the process of the development of a monitoring system. The objectives of monitoring, the foci of the monitoring function, and the organizational setting have been found to be decisive for the final shape of a monitoring system. In monitoring the quality of education, for example, it is important to know for which reason(s) a primary school wants to employ monitoring. Is the main objective assessment of the pupils' achievements over time, the evaluation of the school learning program, or to establish that the financial and other resources contribute to the school's output? Moreover, what exactly is going to be monitored, the achievements of various skills or the pupils' or teachers' individual well-being. Furthermore, can the monitoring system be practically implemented, and finally, which persons in which functions are going to use or will depend on the monitoring information?

These type of questions need to be answered before a monitoring system can be successfully constructed.

The methodological and statistical problems which have been discussed in chapter 2-6 are especially relevant for the refinement and implementation of the monitoring function in the pupil monitoring system LISKAL. Although the first chapter provides general information on monitoring systems the theoretical rationale for the monitoring of the quality of education could be further elaborated (e.g. Scheerens et al., 1988; also Tuijnman & Postlethwaite, 1994).

The study of missing data in panel research has led to a number of findings. First, it has been shown that the 'overall' model implied covariance matrix of the SEM model can be expressed in terms of Kalman smoother state estimates over all subjects in the sample. Not only between pairs of successive time points (diagonal blocks) the covariance matrices can be reproduced by means of the Kalman smoother algorithm, such as in time-series analysis (e.g. Oud, in press), but also between all other pairs (off-diagonal blocks) of time points over which the SEM fit function is to be minimized. This finding is essential for the implementation of the EM algorithm in conjunction with the Kalman smoother within the context of SEM modeling. Second, the SEM fit function was decomposed, which showed the possibility of the M-step of the EM parameter estimation procedure to reduce to an ordinary observed variables structural equation and regression problem.

For several reasons a solution to the problem of missing data is important for a pupil monitoring system to be successful. First, in case individual test scores are missing there is a problem in assessing a pupil's individual achievement level at one or a number of times. Second, in case a monitoring system has to provide norms of reference or baseline information for the assessment of pupils' individual or average (sub)groups achievements over time, it essential that these norms are accurately estimated. Because panel research is typically prone to produce incomplete data and panel attrition can easily lead to biased norm estimates (in case data is not missing completely at random or MCAR), it important to provide answers to these problems. Note that the overlapping cohort design, in a sense, can be seen as an alternative approach to limiting the effects of attrition in repeated measurement designs.

The missing data procedure as proposed in chapter 2 presupposes a causal model of the processes of interest. In fact, the knowledge of these processes is utilized in the missing data procedure. In case of little knowledge less restrictions could be imposed onto the model and on the basis of interpretations of a less restricted model a number of competing models could be estimated and tested for model selection and evaluation. It should be mentioned, however, that in longitudinal research the assumption of the future not influencing the present and past states is a plausible one. From this perspective one could argue that EM as a model based procedure is not as restrictive as is sometimes suggested.

Although the EM algorithm provides a flexible method of solving a rather general problem in social science research, there are still some issues to be solved. First, non-MAR (missing at random) mechanisms cannot be ignored in the EM missing data procedure. More research should be devoted to detecting non-MAR mechanisms in educational data and how these should be handled. Second, the computation of the standard errors is somewhat problematic in EM. In the last M-step standard errors are computed on the basis of the maximum likelihood solution. The problem, however, is what value should be inserted for N in the denominator of the standard errors.

In chapter 3 it has been shown how the EM missing data procedure is adapted as to accommodate for the latent and observed means processes of the structured means SEM model. Also the Kalman filter accommodates for means processes, allowing the estimation of growth curves on an absolute scale. Especially in assessing educational progress by means of LISKAL it is important not only to be able to assess individual growth relatively in comparison to the population mean growth curve, but also to assess progress in terms of absolute growth and decay. Furthermore, by explicitly modeling the latent means over time, and thus making use of raw test scores, in the case of missings, the correct maximum likelihood estimates of the means parameters (latent intercepts and location parameters) are obtained. In contrast, in subtracting the observed means from the raw test scores, deviation scores may become distorted as the observed means are possibly biased because of a 'listwise' or 'pairwise' deletion of cases in the sample. This could, for example, happen in case of a MAR missing data mechanism.

In chapter 4 nonstandard linear and nonlinear constraints have been applied in SEM state space analysis for the modeling of first-order and second-order stationary processes and for the modeling of common means and/or variances-covariances in cohorts of the overlapping cohort design. It has been shown that nonstandard constraints can be employed for constraining both the first- and second-order moments of the latent as well as observed variables processes. The implications of the stationarity constraints, that constrain the latent variables processes, for the time dependence of the model parameters have been clarified. Also the initialization problem, in case no past information is available in the SEM model, is shown to be solvable by means of nonstandard, possibly nonlinear, constraints.

Nonstandard constraints are especially important in the construction of the pupil monitoring system LISKAL. Because LISKAL uses population mean estimates to determine the norms of reference with regard to the contents of the curriculum, and employs this information for the assessment of pupils' achievement levels over time, it is important this information not to be obsolete. This is one of the two major reasons for applying the overlapping cohort design because the data collection period can be considerably shortened. For keeping a pupil monitoring system, that is constantly used in practice, up to date, it has to be regularly fed by new information about the developments in the population of interest. However, in order to obtain the population mean information on the basis of several partly overlapping cohorts, and thus over the entire time zone that is covered by these cohorts, it is important to be able to estimate and to test for commonness in the latent means and latent variances-covariances at the overlapping parts of the overlapping cohort design.

Because SEM applications become more comprehensive it is desirable that SEM programs allow the specification of nonstandard linear and nonlinear constraints by means of matrix algebraic expressions (e.g. in Mx). As has been shown in chapter 4 and also in chapter 6, in which the matrix exponential function is employed as well as higher order polynomial schemes for the modeling of input-effects and stochastic error terms, it allows the use of very complex nonlinear constraints. A complication in the application of nonstandard linear and nonlinear constraints is that it can be a difficult task to determine the exact number of degrees of freedom in the model, which depends on the dimensionality of the maximum likelihood solution.

Chapter 5 relates to previous chapters in the following way. It relates to chapter 1 in that it has been shown how a pupil monitoring system can be constructed from a methodological point of view. Furthermore, some of the results are already implied by the findings in chapter 3, as for example, the formulation and estimation of the structured means SEM model, and filtering and smoothing on the basis of the structured means SEM. However, there is a number of other problems in the construction of LISKAL, that have been postponed until chapter 5.

Because in the behavioral sciences the number of measurement time points is usually small, it is important to be able to initialize the Kalman filter by means of an unbiased estimator. The cross-sectional Bartlett estimator, which has minimum variance in the class of conditionally unbiased estimators, turned out to be an appropriate candidate. In fact, the Bartlett initialized Kalman filter has been proven to be t_0-conditionally unbiased as well as the Kalman smoother, if initialized by means of the t_0-conditionally unbiased Kalman filter. In contrast the minimum variance cross-sectional regression estimator has been shown to be conditionally as well as t_0-conditionally biased. It has been pointed out that the recursive Kalman smoother algorithm is equivalent to the 'overall' regression estimator, but that the latter is numerically less efficient. Finally, an alternative unbiased estimator has been derived for the initialization of the Kalman filter in case trait variables are added to the state vector.

The construction of a monitoring system by means of the SEM state space approach, the estimation of the population mean development and of the individual state trajectories have been treated for the structured means as well as for the state-trait model. The state trajectories in the state-trait model have been shown to regress towards (or to egress from) the subject specific zero-initial-mean instead of the population mean as is usually the case in longitudinal SEM models. Two approaches that guarantee the latent measurement scales to maintain the same origins and measurement units across the entire time zone of the SEM model have been explained. These contributions emphasize the importance of the choice of the measurement instruments; first to be able to construct the latent variables, and second to be able to assess educational growth in terms of well-defined latent scales.

In chapter 6 it has been clarified that for a wide range of state space models, that are translatable into SEM, continuous time versions can also be estimated by means of a suitable SEM program. The continuous time version of the discrete time state-trait model has been shown to be estimable by means of SEM. The constant subject specific effects and the continuous time state trajectories are estimable on the basis the exact discrete model by application of the discrete time (over arbitrary time intervals) Kalman filter and Kalman smoother. In addition to deriving the exact discrete model under the condition of time-invariance, it has also been derived for parameter matrices being piecewise time-invariant or continuously time-varying according to a polynomial scheme. The matrix exponential function for varying Δt can be implemented for SEM programs that employ general matrix algebraic constraints such as Mx.

Although there are several compelling reasons for the use of continuous time models it has special relevance for the monitoring of educational progress. While in discrete time modeling measurements must be taken at the specific discrete time points as defined in the model, teachers often are not able to comply with the prescribed sampling schemes, possibly because of other priorities as regards the school learning program (see chapter 1 on the practical implementability of monitoring systems). As a result pupils' test scores often cannot be assigned to these prescribed discrete time points. Continuous time sampling makes the time points at which tests are taken by teachers virtually arbitrary.

It has been shown that irregular sampling can be employed in continuous time modeling by means of the SEM state space approach. Thus, the spacing of observation time points may be irregular over time within subjects but, in model estimation, is assumed to be the same between subjects in the sample. As this does not cover the case of different time sampling schemes for different subjects, an alternative approach is suggested below. It makes use of the EM missing data procedure.

The EM missing data procedure is applicable for the exact discrete model as well. By adapting the E-step of the EM procedure different time sampling schemes of different subjects can additionally be handled within the SEM state space model. The E-step is adapted as follows.

First, in computing the Kalman filter and smoother estimates the exact discrete model is adapted by varying the time interval Δt to match the true observation time points of each specific subject. For measurements at time points t'_i preceding the planned observation time point t_i, Δt is reduced accordingly to $\Delta t = t'_i - t_{i-1}$

and for measurements at later points in time it is increased accordingly, so that in both cases each true observation time point t'_i on the continuous time scale is matched exactly by the exact discrete model. Next, a second, corrective exact discrete model is applied with $\Delta t = t_i - t'_i$, going from the true observation time point t'_i to the planned observation time point t_i (where the data is missing), to compute predictive Kalman filter and smoother estimates at t_i. For missing data (no true observation time point at all in the planned interval for the specific subject), only one exact discrete model is used with $\Delta t = t_i - t_{i-1}$, going directly from the preceding planned observation time point to the next planned observation time point. As all available information is processed in each iteration of the EM algorithm and the corrections toward the planned observation time points are the conditional expectations on the basis of the continuous time model estimated, the results converge again to the maximum likelihood estimate of the continuous time model.

A problem still to be solved in continuous time modeling is how the drift matrix $\mathbf{A}(t)$, which was assumed to be (piece-wise) time-invariant, can be modeled in continuous time in a time-varying way. A numerical solution may be necessary. As in the behavioral sciences the hypothesis of autoregression effects varying over time seems a plausible one, it is important to be able to generalize this hypothesis analogously to continuous time.

References

Aarnoutse, C.A.J., van Leeuwe, J.F.J., Oud, J.H.L., Voeten, M.J.M., & van Kan, N.P.L.M. (1996a). *Handleiding Nijmeegs leerlingvolgsysteem: LISKAL* [User's guide of the Nijmegen pupil monitoring system LISKAL]. Manuscript in preparation.

Aarnoutse, C.A.J., van Leeuwe, J.F.J., Voeten, M.J.M., van Kan, N.P.L.M., & Oud, J.H.L. (1996b). *Longitudinaal onderzoek schoolvorderingen in het basisonderwijs* [Longitudinal research of educational progress in primary school education]. Nijmegen: ITS.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control, 19,* 716-723.

ARBO, (1988). *Voorrang aan achterstand: Advies over een integraal beleid ter voorkoming en bestrijding van onderwijsachterstand* [Priority to the underprivileged: Advice about an integral policy to prevent and control educational lags]. Zeist: ARBO Onderwijscentrum.

Arminger, G. (1986). Linear stochastic differential equations for panel data with unobserved variables. In N.B. Tuma (Ed.), *Sociological methodology* (pp. 187-212). Washington: Jossey-Bass.

Arnold, L. (1974). *Stochastic differential equations.* New York: Wiley.

Baltagi, B.H. (1995). *Econometric analysis of panel data.* Chichester: Wiley.

Bell, R.Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24,* 145-152.

Bell, R.Q. (1954). An experimental test of the accelerated longitudinal approach. *Child Development, 25,* 281-286.

Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling *Annual Review of Psychology, 31,* 419-456.

Bentler, P.M., & Jamshidian, M. (1994). Gramian matrices in covariance structure modeling. *Applied Psychological Measurement, 18,* 79-94.

Bergstrom, A.R. (1984). Continuous time stochastic models and issues of aggregation over time In Z. Griliches & M.D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 1145-1212). Amsterdam: North-Holland.

Bergstrom, A.R. (1988). The history of continuous-time econometric models. *Econometric Theory, 4,* 365-383.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K.A., & Long, J. Scott. (1993). *Testing structural equation models.* Newbury Park, California: Sage Publications, Inc.

Boomsma, A. (1983). *On robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.

Brus, B Th., & Voeten, M J.M. (1979). *Eén-Minuut-Test, vorm A en B.* Nijmegen: Berkhout.

Caines, P.E. (1988). *Linear stochastic systems.* New York: Wiley.

Cardon, L.R., Fulker, D.W., & Jöreskog, K.G. (1991). A LISREL 8 model with constrained parameters for twin and adoptive families. *Behavior Genetics, 21,* 327-350.

Casley, D.J., & Kumar, K. (1987). *Project monitoring and evaluation in agriculture.* Baltimore, Maryland: The John Hopkins University Press.

Casley, D.J., & Kumar, K. (1988). *The collection, analysis, and use of monitoring and evaluation data.* Baltimore, Maryland: The John Hopkins University Press.

Coleman, J.S. (1968). The mathematical study of change. In H.M. Blalock, Jr. & A.B. Blalock (Eds.), *Methodology in social research* (pp. 428-478). New York: McGraw-Hill.

Dembo, A., & Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications, 23,* 91-113.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1-38.

Desoer, C.A. (1970). *Notes for a second course on linear systems.* New York: Van Nostrand Reinhold.

Dolan, C.V., & Molenaar, P.C.M. (1993). NONLIS: A FORTRAN program for multi-group covariance structure analysis with non-standard constraints. In F.J. Maarse, A.E. Akkerman, A.N. Brand, L.J.M. Mulder & M.J. van der Stelt (Eds.), *Computers in psychology: Tools for experimental and applied psychology* (pp. 105-115). Amsterdam: Swets and Zeitlinger.

Duncan, S.C., Duncan, T.E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods, 1,* 236-248.

Eggen, T., Engelen, R., & Kamphuis, F. (1991). *Methodological aspects of the student monitoring system for primary schools.* Unpublished manuscript.

Gillijns, P. (1991). *Leerlingvolgsysteem* [Pupil monitoring system]. Tilburg: Zwijsen.

Gillijns, P. (1994). *Computerprogramma leerlingvolgsysteem: Handleiding* [Computerprogram pupil monitoring system: User's guide]. Arhnem: Cito.

Goossens, M., Mittelbach, F., & Samarin, A. (1994). *The LaTeXCompanion.* New York: Addison-Wesley.

Gosovic, B. (1992). *The quest for world environmental cooperation: The case of the UN Global Environment Monitoring System.* London: Routledge Press.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory Principles and applications.* Dordrecht, The Netherlands: Kluwer · Nijhoff Publishing.

Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology, 16,* 201-220.

Hamerle, A., Singer, H., & Nagl, W. (1993). Identification and estimation of continuous time dynamic systems with exogenous variables using panel data. *Econometric Theory, 9,* 283-295.

Hansen, L.P., & Sargent, T.J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica, 51,* 377-387.

Haughton, D.M.A., Oud, J.H.L., & Jansen, R.A.R.G. (in press). Information and other criteria in structural equation model selection. *Communications in Statistics, Part B.*

Horn. J.J., & McArdle, J.J. (1980). Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L.W. Poon (Ed.), *Aging in the 1980's: Psychological issues* (pp. 503-541). Washington, DC: APA.

Hsiao, C. (1986). *Analysis of panel data.* New York: Cambridge University Press.

Husén, T., & Tuijnman, A. (1994) Monitoring standards in education: Why and how it came about. In A.C. Tuijnman & T.N. Postlethwaite (Eds.), *Monitoring the standards of education: Papers in honor of John P. Keeves* (pp. 1-21). Oxford: Pergamon Press.

Jansen, R.A.R.G., & Oud, J.H.L. (1995). Handling missing data in the construction of the pupil monitoring system LISKAL. In I. Partchev (Ed.), *Proceedings of the SMABS 1994 conference: Multivariate analysis in the behavioral sciences: Philosophic to technical* (pp. 39-48). Sofia: Prof. Marin Drinov Publishing House.

Jansen, R.A.R.G., & Oud, J.H.L. (1995). Longitudinal LISREL model estimation from incomplete data using the EM algorithm and the Kalman smoother. *Statistica Neerlandica, 49,* 362-377.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443-482.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183-202.

Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-107). New York: Seminar Press.

Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1-56). San Francisco CA: Freeman.

Jöreskog, K.G. (1978). An econometric model for multivariate panel data. *Annales de l'INSEE, 30-31,* 355-366.

Jöreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage Publications, Inc.

Jöreskog, K.G., & Sörbom, D. (1985). Simultaneous analysis of longitudinal data from several cohorts. In W.M. Mason & S.E. Fienberg (Eds.), *Cohort analysis* *in social research: Beyond the identification problem* (pp. 323-341). New York: Springer.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL7: User's reference guide.* Mooresville IN: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1993). *New features in LISREL 8.* Chicago: Scientific Software International.

Jazwinski, A.H. (1970). *Stochastic processes and filtering theory.* New York: Academic Press.

Jencks, S.F. (1994). HCFA's Health Care Quality Improvement Program and the cooperative cardiovascular project. *Annual Thorac Surgery, 58,* 1858-1862.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Transactions of the ASME, Series. D.), 82,* 35-45.

Kalman, R.E., & Bucy R.S. (1961). New results in linear filtering and prediction problems. *Journal of Basic Engineering (Transactions of the ASME, Series D.), 83,* 95-108.

Kamphuis, F.H. (1993). Estimation and prediction of individual ability in longitudinal studies. In J.H.L. Oud & R.A.W. van Blokland-Vogelesang (Eds.), *Proceedings of the SMABS 1992 conference: Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 41-52). Nijmegen: ITS.

Lawley, D.N., & Maxwell, A.E. (1971). *Factor analysis as a statistical method.* London: Butterworths.

Lewis, F.L. (1986). *Optimal estimation: With an introduction to stochastic control theory.* New York: John Wiley & Sons, Inc.

Liptser, R.S., & Shiryayev, A.N. (1977, 1978). *Statistics of random processes* (Vols. 1-2). New York: Springer.

Little, R.J.A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association, 87,* 1227-1237.

Little, D.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Luenberger. D.G. (1979). *Introduction to dynamic systems: Theory, models and applications.* New York: Wiley.

MacCallum, R., & Ashby, F.G. (1986). Relationships between linear system theory and covariance structure modeling. *Journal of Mathematical Psychology, 30,* 1-27.

Marangoni, S. (1994). Cardiological monitoring in a respiratory intermediate intensive care unit. *Monaldi Archives for Chest Disease, 49,* 504-507.

McArdle, J.J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L.M. Collins & J.L. Horn (Eds.), *Best methods for the analysis of change: Recent ad-*

vances, unanswered questions, future directions (pp. 276-298). Washington, DC: American Psychological Association.

Meditch, J.S. (1969). *Stochastic optimal linear estimation and control.* New York: McGraw-Hill.

Melis, G., & Sonsma, F. (1989). *Het SAVU-leerlingvolgsysteem: Functie-opzet-uitvoering* [The SAVU pupil monitoring system: Function-design-implementation]. Hoevelaken: Christelijk Pedagogisch Studiecentrum.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107-122.

Moelands, F., Mommers, C., & Oud, H (1990). Leerlingvolgsystemen verklaard en vergeleken [Pupil monitoring systems explained and compared]. *School & Begeleiding, 7,* 19-28.

Molenaar, J., & Oud, J.H.L. (1991). Optimality and initialization of the Kalman filter. *Kwantatieve Methoden, 12*(38), 45-52.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171-189.

Neale, M.C. (1995). *Mx: Statistical Modeling* (3rd ed.). Richmond: Medical College of Virginia, Virginia Commonwealth University, Department of Psychiatry.

Otter, P.W. (1984). *Dynamic feature space modelling, filtering and self-tuning control of stochastic systems: A systems approach with economic and social applications.* Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.

Oud, J.H.L. (1978) . *Systeem-methodologie in sociaal-wetenschappelijk onderzoek* [Systems methodology in social science research]. Nijmegen: Alfa.

Oud, J.H.L. (in press). SEM state space modeling of panel data and its relationship to traditional single subject state space modeling. *Kwantitatieve Methoden.*

Oud, J.H.L., van den Bercken, J.H.L., & Essers, R.J. (1990). Longitudinal factor scores estimation using the Kalman filter. *Applied Psychological Measurement, 14,* 395-418.

Oud, J.H.L., Haughton, D.M.A., & Jansen, R.A.R.G. (1996). Information and other criteria in structural equation model selection. *Kwantitatieve Methoden, 17*(52), 69-100.

Oud, J.H.L., & Jansen, R.A.R.G. (1995) An ARMA extension of the longitudinal LISREL model for LISKAL. In I. Partchev (Ed.), *Proceedings of the SMABS 1994 conference: Multivariate analysis in the behavioral sciences: Philosophic to technical* (pp. 46-69). Sofia: Prof. Marin Drinov Publishing House.

Oud, J.H.L., & Jansen, R.A.R G. (1996). Nonstationary longitudinal LISREL model estimation from incomplete panel data using EM and the Kalman smoother. In U. Engel & J Reinecke (Eds.), *Analysis of change: Advanced techniques in* *panel data analysis* (pp. 135-159). New York: de Gruyter.

Oud, J.H.L., & Mommers, M.J C. (1990). De valkuil van het didactische leeftijdsequivalent [The pitfall of educational age]. *Tijdschrift voor Orthopedagogiek, 19,* 445-459.

Oud, J.H.L., Mommers, M.J.C., & Heijmans, M. (1991). *LISKAL-M1: Een leerlingvolgsysteem voor groep 3, 4 en 5 van de basisschool* [LISKAL-M1: A pupil monitoring system for grade 1, 2, and 3 of primary school]. Nijmegen: Instituut voor Orthopedagogiek.

Oud, J.H.L., Mommers, M.J.C., Smitshuis, A., Doppenberg, E., Devilee, G., & Heijmans, M. (1993). *LISKAL-M2: Een leerlingvolgsysteem voor groep 3, 4 en 5 van de basisschool* [LISKAL-M2: A pupil monitoring system for grade 3, 4, and 5 of primary school]. Nijmegen: Instituut voor Orthopedagogiek.

Oud, J.H.L., van Leeuwe, J.F.J., & Jansen, R.A.R.G. (1993). Kalman filtering in discrete and continuous time based on longitudinal SEM models. In J.H.L. Oud & R.A.W. van Blokland-Vogelesang (Eds.), *Proceedings of the SMABS 1992 conference: Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 3-26). Nijmegen: ITS.

Pelgrum, W.J. (1990). *Educational assessment: Monitoring, evaluation and the curriculum.* Unpublished doctoral dissertation, University of Twente, The Netherlands.

Phillips, P.C.B. (1973). The problem of identification in finite parameter continuous time models. In A.R. Bergstrom (Ed.), *Statistical inference in continuous time models* (pp. 135-173). Amsterdam: North-Holland.

Plomp, T., Huijsman, H., & Kluyfhout, E. (1992). Monitoring in educational developmental projects: The development of a monitoring system. *International Journal of Educational Development, 12,* 65-73.

Priestly, M.B. (1981). *Spectral analysis and time series: Vol. 2. Multivariate series, prediction and control.* London: Academic Press.

Putte, R.A. van de, (1991). *Monitoring for development control: The design of a model.* Unpublished doctoral dissertation, Twente, The Netherlands.

Rauch, H.E., Tung, F., & Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal, 3,* 1445-1450.

Raudenbush, S.W., & Chan, W.S. (1992). Growth curve analysis in accelerated longitudinal designs. *Journal of Research in Crime and Delinquency, 29,* 387-411.

Raudenbush, S.W., & Chan, W. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61,* 941-951.

Rindskopf, D.M. (1984). Using phantom and imaginary latent variables to parametrize constraints in linear structural models. *Psychometrika, 49,* 37-47.

Rogosa, D.R., & Willet, J.B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203-228.

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63,* 581-592.

Ruymgaart, P.A., & Soong, T.T. (1985). *Mathematics of Kalman-Bucy filtering.* Berlin: Springer.

Scheerens, J. Stoel, W.G.R., Vermeulen, C.J.A.J., & Pelgrum, W.J. (1988). *De haalbaarheid van een onderwijsindicatoren stelsel voor het basis- en voortgezet onderwijs* [The feasibility of a system of educational indicators for elementary and secondary education]. Enschede: University of Twente, Center for Applied Educational Research.

Shumway, R.H., & Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis, 3,* 253-264.

Singer, H. (1990). *Parameterschätzung in zeitkontinuierlichen dynamischen Systemen* [Parameter estimation in continuous time dynamic systems]. Konstanz: Hartung-Gorre.

Singer, H. (1991). LSDE - a program for the simulation, graphical display, optimal filtering and maximum likelihood estimation of linear stochastic differential equations: User's guide. Meersburg, FRG: Author.

Singer, H. (1992). *Zeitkontinuierliche dynamischen Systeme* [Continuous time dynamical systems]. Frankfurt a.M.: Campus.

Singer, H. (1993). Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components. *Journal of Time Series Analysis, 14,* 527-545.

Tilburg, P. van, & de Haan (1995). *Controlling development: Systems of monitoring & evaluation and management information for project planning in developing countries.* (pp. 86-107). Tilburg: Tilburg University Press.

Tilburg, P. van. de Haan, J., & Giesberts, A. (1995). A system for monitoring & evaluation and management information (MEMIS): An integrated system for planning purposes in developing countries. In P. van Tilburg & J. de Haan (Eds.), *Controlling development: Systems of monitoring & evaluation and management information for project planning in developing countries* (pp. 86-107). Tilburg: Tilburg University Press.

Traub, R.E. (1983). A priori considerations in choosing an item response model. In R.K. Hambleton (Ed.), *Applications of item response theory* (pp. 57-70). Vancouver, B.C.: Educational Research Institute of British Columbia.

Tuijnman, A.C., & Postlethwaite, T.N. (1994). *Monitoring the standards of education: Papers in honor of John P. Keeves.* Oxford: Pergamon Press.

Tuma, N.B., & Hannan, M.T. (1984). *Social dynamics.* New York: Academic Press.

Vitacca, M., & Clini, E. (1994). Respiratory monitoring in an intermediate intensive unit. *Monaldi Archives for Chest Disease, 49,* 508-512.

Wechsler, D (1974). *Manual for the Wechsler intelligence scale for children - revised.* New York: The Psychological Corporation.

Willems, J.C. (1991). Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control, 36,* 259-295.

Willett, J.B., & Sayer, A.G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363-381.

Wold, H.O.A. (1954). Causality and econometrics. *Econometrica, 40,* 162-177.

Zadeh, L.A., & Desoer, C.A. (1963). *Linear system theory: The state space approach.* New York: McGraw-Hill.

Samenvatting

Constructie van Volgsystemen in de Gedragswetenschappen: De SEM Toestand Ruimte Benadering

Het doel van het proefschrift is een bijdrage te leveren aan de ontwikkeling van volgsystemen op het terrein van het onderwijs. In hoofdstuk 1 worden de belangrijkste onderdelen van 'monitoring' (volgen) en 'monitoring systems' (volgsystemen) besproken. De methodologische en statistische problemen komen aan de orde in hoofdstukken 2-6. Hoofdstuk 7, tenslotte, bestaat uit een opsomming en een discussie van de onderzoeksresultaten in termen van implicaties voor de constructie van volgsystemen in het basisonderwijs.

In *hoofdstuk 1* wordt een inleiding gegeven over volgsystemen in het algemeen. Vanwege de grote variëteit van volgsystemen in de praktijk wordt het onderwerp in een breed kader geplaatst. Aan de orde komen hoofdonderdelen, doelstellingen en andere aspecten van volgsystemen. Vervolgens worden volgsystemen besproken binnen de context van het onderwijs, ofwel het vaststellen van onderwijsvorderingen. Naast het SEM[1] toestand ruimte model, dat centraal staat in het onderzoek, worden twee andere benaderingen voor het volgen van onderwijsvorderingen in het basisonderwijs besproken. Het eerste hoofdstuk draagt op twee manieren bij aan het centrale thema van het proefschrift. Het geeft algemene achtergrond informatie over 'monitoring' en 'monitoring systems', en het stelt de relevante context vast van de methodologische en statistische problemen die onderwerp zijn van hoofdstukken 2-6.

In *hoofdstuk 2* wordt het probleem 'missing data' (ontbrekende gegevens) in panel data sets behandeld. Het ontbreken van gegevens is een algemeen probleem in sociaal wetenschappelijk onderzoek maar doet zich met name voor in panel onderzoek. 'Missing data' impliceren het verlies van informatie, namelijk wanneer statistische analyses beperkt worden tot de groep van onderzoekseenheden waarvan geen gegevens ontbreken. Na een aantal herhaalde metingen kan panel 'attrition' (uitval) ertoe leiden dat de omvang van de steekproef afneemt tot een fractie van de oorspronkelijke omvang. Het ontbreken van gegevens is vaak gerelateerd aan de scores op de gemeten variabelen. Bijvoorbeeld, leerlingen die een laag niveau hebben, hebben een grotere kans op uitval juist vanwege het lage niveau. Als gevolg hiervan kunnen gegevens systematisch ontbreken, hetgeen tot vertekening kan leiden in populatie parameter schattingen. Er worden verschillende 'missing data' mechanismen onderscheiden, die, afhankelijk van de voorgestelde 'missing data' procedure, genegeerd dan wel niet genegeerd kunnen worden.

De 'missing data' procedure die wordt voorgesteld voor de SEM analyse van panel data sets, maakt gebruik van het EM (Expectation-Maximization) algoritme

[1] SEM staat voor 'Structurele Vergelijkingen Model'.

gecombineerd met de Kalman smoother ter berekening van maximum likelihood schattingen van het longitudinale SEM model op basis van variërende missing data patronen. Het is een modelgebaseerde procedure die bepaalde relaties veronderstelt tussen de variabelen die bestudeerd worden. Het EM algoritme staat toe dat de data 'missing at random' (MAR) zijn, en heeft een aantal voordelen vergeleken met zogenaamde 'ad-hoc' procedures.

Hoofdstuk 3 geeft een uitbreiding aan het longitudinale SEM model in de behandeling van ontbrekende gegevens. Getoond wordt hoe de 'missing data' procedure kan worden toegepast in de constructie van het leerlingvolgsysteem LISKAL. In plaats van het 'zero means' wordt het 'structured means' SEM model gedefinieerd, hetgeen impliceert dat het SSM[2] geformuleerd dient te worden in termen van latente en geobserveerde gemiddelden processen. Bijgevolg kunnen individuele latente ontwikkelingscurven worden geschat op een absolute schaal. Het niveau van een leerling op één tijdstip kan worden vergeleken met eerdere of latere tijdstippen in termen van absolute groei of afname in groei. In hoofdstuk 3 wordt de term SEM gebruikt in plaats van LISREL omdat de 'missing data' procedure ook kan worden toegepast met behulp van een ander SEM programma dan LISREL.

In *hoofdstuk 4* wordt aangetoond hoe algemene niet-standaard lineaire en nonlineaire restricties kunnen worden toegepast in longitudinale SEM modellen. Niet-standaard restricties worden toegepast voor het modelleren van eerste- en tweede-orde stationaire processen. Algemene matrix algebraïshe expressies worden afgeleid en toegepast om de eerste- en tweede-orde momenten van de latente en geobserveerde variabelen te restricteren. De implicaties van stationariteit voor de tijdsafhankelijkheid van de model parameters worden toegelicht. Niet-standaard restricties worden toegepast voor het modelleren van groei op basis van het overlappende cohort model (OCD) ook wel het geaccelereerde longitudinale design genoemd. Een dergelijke aanpak is van belang voor een efficiënte constructie van volgsystemen en is uitgebreid toegepast voor de constructie van LISKAL. Met behulp van niet-standaard restricties kan worden getest of partieel overlappende cohorten gemeenschappelijke model geïmpliceerde karakteristieken hebben in de vorm van latente gemiddelde trajecten en latente covariantie functies, die achtereenvolgens kunnen worden geschat en worden gebruikt in een volgsysteem. Op basis van gemeenschappelijke latente gemiddelden en latente varianties-covarianties kunnen ontwikkelingscurven worden bepaald die de gehele relevante tijdsperiode bestrijken.

In *hoofdstuk 5* worden verschillende bijzondere gevallen van het basis SSM besproken, en er wordt aangetoond hoe deze modellen vertaald kunnen worden in SEM. De Kalman filter en Kalman smoother worden toegepast voor het 'structured means' SEM en het 'state-trait' model. De SEM formulering en parameter

[2] SSM staat voor 'Toestand Ruimte Model'

schatting van het SSM met input-effecten worden in detail besproken. Er wordt een vergelijking gemaakt tussen het Kalman filter en twee bekende cross-sectionele factorscore schatters. Een belangrijke vraag richt zich op de keuze van één van de cross-sectionele schatters voor de initialisatie van het Kalman filter en heeft betrekking op de criteria van minimale variantie en zuiverheid van een schatter. Er wordt gewezen op de relatie tussen de 'overall' regressieschatter en de Kalman smoother, alsmede t_0-conditionele zuiverheid van de smoother. Tenslotte, wordt getoond hoe het probleem van de initialisatie van het Kalman filter voor het 'state-trait' model kan worden opgelost met behulp van de Bartlett schatter.

In *hoofdstuk 6* wordt ingegaan op de vraag hoe met behulp van het SEM maximum likelihood schattingen verkregen kunnen worden van het lineaire stochastische continu tijd SSM. Het zogenaamde exact discrete model (EDM) wordt afgeleid onder vrij algemene condities: parameter matrices mogen variëren volgens een stuksgewijs tijd-invariant schema, dan wel volgens een continu tijdvariërend polynoom schema. Aangezien de parameters van het EDM complexe nonlineaire functies zijn van de parameters van het continu tijd systeem, dienen complexe nonlineaire relaties opgelegd te worden tijdens de SEM schattingsprocedure. Deze aanpak, die bekend staat als de 'directe methode', vormt een alternatief voor de sterk bekritiseerde 'indirecte methode', die in het verleden in SEM toepassingen gebruikt werd.

Tenslotte worden in *hoofdstuk 7* de belangrijkste bevindingen opgesomd. De relaties tussen de inhoud van de verschillende hoofdstukken worden benadrukt. Speciale aandacht gaat uit naar de relevantie van de verschillende bijdragen voor de constructie van volgsystemen op het terrein van het basisonderwijs. Een aantal resultaten wordt van kritische commentaren voorzien en er wordt een aantal suggesties gedaan voor toekomstig onderzoek op dit terrein.

Curriculum vitae

Robert Jansen, geboren op 31 januari 1968, studeerde na het behalen van het VWO-diploma in 1986, sociologie aan de Katholieke Universiteit Nijmegen. Het doctoraalexamen werd in 1992 afgelegd. Van 1992 tot 1996 was hij werkzaam als AIO binnen het onderzoekprogramma Psychodiagnostiek, dat deel uitmaakt van de Nijmeegse Universitaire Onderzoekschool voor Opvoeding en Onderwijs (NUOVO), alwaar hij methodologisch onderzoek verrichtte naar constructie van volgsystemen voor sociaal-wetenschappelijke toepassingen.

Robert Jansen was born on January the 31st 1968, and studied sociology at the University of Nijmegen after obtaining his pre-university educational certificate in 1992. Between 1992 and 1996 he was a PhD student in the research program 'Psychological Assessment' (Psychodiagnostiek), which is part of the Nijmegen University Research School (NUOVO), where he performed methodological research on constructing monitoring systems for applications within the field of the social sciences.